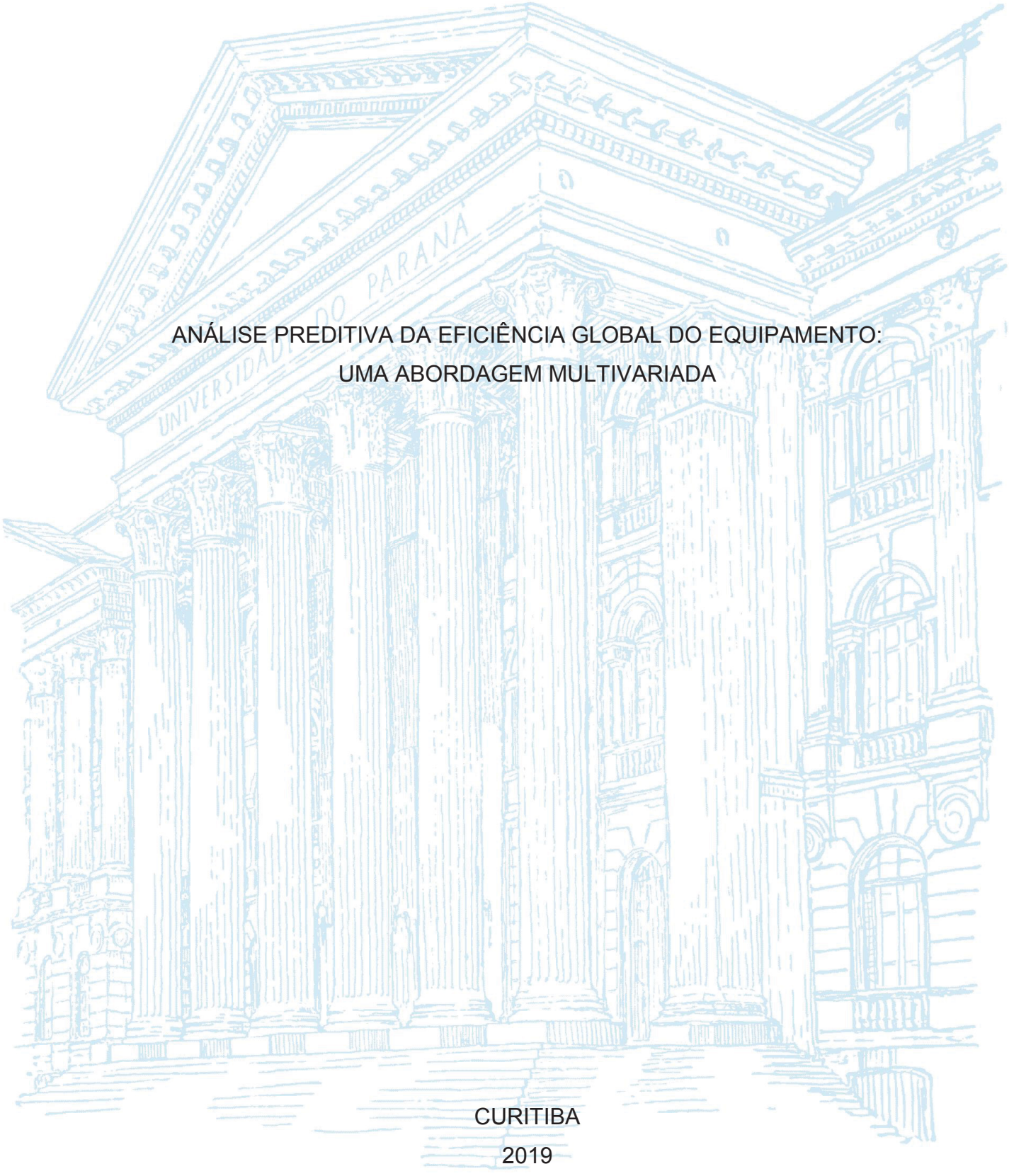
UNIVERSIDADE FEDERAL DO PARANÁ

DANIEL AYUB



ANÁLISE PREDITIVA DA EFICIÊNCIA GLOBAL DO EQUIPAMENTO: UMA ABORDAGEM MULTIVARIADA

CURITIBA

2019

DANIEL AYUB

ANÁLISE PREDITIVA DA EFICIÊNCIA GLOBAL DO EQUIPAMENTO: UMA ABORDAGEM MULTIVARIADA

Dissertação apresentada ao programa de Pós-Graduação em Engenharia de Produção, Setor de Tecnologia, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção.

Orientador: Prof. Dr. Marcos Augusto Mendes Marques

Coorientadora: Profa. Dra. Mariana Kleina

CURITIBA

2019

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

A988a

Ayub, Daniel
Análise preditiva da eficiência global do equipamento: uma abordagem multivariada / Daniel Ayub. – Curitiba, 2019.

Dissertação - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Produção, 2019.

Orientador: Marcos Augusto Mendes Marques. Coorientador: Mariana Kleina.

1. Análise multivariada. 2. Redes neurais (Computação). 3. Big data. 4. Manutenção produtiva total. I. Universidade Federal do Paraná. II. Marques, Marcos Augusto Mendes. III. Kleina, Mariana IV. Título.

CDD: 331.04

Bibliotecária: Vanusa Maciel CRB- 9/1928

MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA DE PRODUÇÃO - 40001016070P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA DE PRODUÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **DANIEL AYUB** intitulada: **ANÁLISE PREDITIVA DA EFICIÊNCIA GLOBAL DO EQUIPAMENTO: UMA ABORDAGEM MULTIVARIADA**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 17 de Junho de 2019.

MARCOS AUGUSTO MENDES MARQUES
Presidente da Banca Examinadora

ROBSON SELEME
Avaliador Interno (UFPR)

MARCELO GECHELE CLETO
Avaliador Interno (UFPR)

VOLMIR EUGÊNIO WILHELM
Avaliador Externo (UFPR)

ISABELLA ANDRECZEVSKI CHAVES
Avaliador Externo (UFPR)

Dedico primeiramente a Deus que está sempre conosco,
aos meus mestres e mentores,
por todos os ensinamentos e inspirações.
A minha querida família, que esteve sempre ao meu lado,
tendo carinho e compreensão em todos os momentos,
principalmente aos meus maravilhosos filhos Wesley e Giovana,
e a minha mãe Aldeli que sempre me incentivaram em tudo.

“Para ser superado, não é preciso parar, basta permanecer com a mesma velocidade, atitude ou forma de pensar”.

(*Eliyahu Goldratt - A Meta, 1984*)

“No que diz respeito ao desempenho, ao compromisso, ao esforço, à dedicação, não existe meio termo, ou você faz uma coisa bem feita, ou não faz.”

(*Ayrton Senna*)

RESUMO

As organizações visando o foco na manutenção de sua competitividade, buscam continuamente a gestão e melhoria de seus processos. Uma das metodologias muito utilizadas visando a redução de perdas e desperdícios é denominada de Manutenção Produtiva Total (TPM), a qual contempla técnicas para auxiliar na melhoria do desempenho de um processo, sendo uma delas conhecida como Eficiência Global do Equipamento (OEE), a qual permite demonstrar como a produção está se comportando, revelando desvios, desperdícios e perdas do processo, gerando informações para análise de dados. Com a aplicação dos conceitos da Indústria 4.0, mais especificamente em relação a utilização de equipamentos de coleta, transmissão e armazenamento de dados cada vez mais eficientes, observa-se uma quantidade expressiva de informações geradas a partir dos processos de fabricação. Tal realidade permite acesso a uma quantidade considerável de dados para a análise do OEE, a qual traz históricos sobre o comportamento dos equipamentos, processos, produtos, paradas de máquina, defeitos, entre outros. No presente trabalho, estas informações passaram por métodos analíticos multivariados, os quais objetivaram o retorno de previsões acerca da eficiência dos equipamentos. Neste processo, buscando a geração de um modelo preditivo para a análise do OEE, foram aplicadas técnicas estatísticas e multivariadas, como a análise da matriz de correlação, componentes principais e regressão linear múltipla. Porém, os resultados de maior relevância foram originados pela aplicação da técnica de Redes Neurais Artificiais (RNA), as quais retornaram previsões com boa exatidão sobre o OEE em relação as variáveis do processo, além de fornecer informações para a construção de elementos gráficos, os quais proporcionaram a observação do comportamento da eficiência do equipamento analisado em relação as variáveis preditoras. Estas analises, em alguns casos, trouxeram confirmações a respeito de convicções empíricas do processo, demonstrando como certas perdas do equipamento relacionadas a disponibilidade, performance e qualidade, impactam de forma positiva ou negativa nos índices de OEE. Tais resultados preditivos podem fornecer conjunturas de apoio para tomada de decisões precisas e oportunas, e com possíveis alternativas para roteiros mais eficientes e de menores custos. Neste âmbito, a análise de dados por meio de Redes Neurais Artificiais demonstrou que pode viabilizar resultados consideráveis para um ambiente de grande quantidade de dados e com uma expressiva variabilidade de informações. A análise de dados é cada vez mais reconhecida como um valioso conjunto de técnicas para aperfeiçoar o desempenho das empresas, desta forma, o intuito deste trabalho é demonstrar uma alternativa para a análise preditiva de OEE, por meio de abordagens multivariadas aplicadas em ambientes providos de grandes quantidades de dados.

Palavras-chave: OEE. Análise preditiva. Análise multivariada. Redes neurais artificiais. Indústria 4.0. *Big Data*.

ABSTRACT

The organizations aiming at the maintenance of their competitiveness, continuously seek the management and improvement of their processes. One of the widely used methodologies used to reduce losses and wastes is called Total Productive Maintenance (TPM), which includes techniques to help improve the performance of a process, one of which is known as Global Equipment Effectiveness (OEE). which allows to demonstrate how the production is behaving, revealing deviations, wastes and losses of the process, generating information for data analysis. With the application of the concepts of Industry 4.0, more specifically in relation to the use of equipment of collection, transmission and storage of data increasingly efficient, an expressive amount of information generated from the manufacturing processes is observed. This reality allows access to a considerable amount of data for the OEE analysis, which brings history about the behavior of equipment, processes, products, machine stops, defects, among others. In the present work, this information went through multivariate analytical methods, which aimed to return predictions about the efficiency of the equipment. In this process, statistical and multivariate techniques, such as correlation matrix analysis, main components and multiple linear regression, were used to generate a predictive model for OEE analysis. However, the results of greater relevance were originated by the application of the technique of Artificial Neural Networks (RNA), which returned predictions with good accuracy on the OEE in relation to the process variables, besides providing information for the construction of graphic elements, the which provided the observation of the behavior of the efficiency of the analyzed equipment in relation to the predictive variables. These analyzes have in some cases brought confirmation of the empirical convictions of the process, demonstrating how certain equipment losses related to availability, performance and quality have a positive or negative impact on the OEE indexes. Such predictive results can provide support scenarios for accurate and timely decision making, and possible alternatives to more cost-effective and cost-effective roadmaps. In this context, data analysis through Artificial Neural Networks has demonstrated that it can provide considerable results for an environment with a large amount of data and with an expressive variability of information. Data analysis is increasingly recognized as a valuable set of techniques to improve the performance of companies, thus, the purpose of this work is to demonstrate an alternative to the predictive analysis of OEE, through multivariate approaches applied in environments provided by large amounts of data.

Keywords: OEE. Predictive analysis. Multivariate analysis. Artificial neural networks. Industry 4.0. Big Data.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ACP	-	Análise de Componentes Principais
BDA	-	Análise de *Big Data* (*Big Data Analytics*)
CAPES	-	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CLP	-	Controlador lógico programável
CPS	-	Sistema Cyber-Físico (*Cyber-Physical System*)
DOC.	-	Documento
EBS	-	Sistema de Negócio Empresarial (*Enterprise Business Systems*)
EUA	-	Estados Unidos da América
IHM	-	Interface Homem-Máquina
IoT	-	Internet das Coisas (*Internet of Things*)
KPI	-	Indicador-Chave de Desempenho (*Key Performance Indicator*)
OEE	-	Eficiência Global do Equipamento (*Overall Equipment Effectiveness*)
PCP	-	Planejamento e Controle da Produção
RFID	-	Identificador por Radiofrequência (*Radio-Frequency IDentification*)
SQL	-	Linguagem de Consulta Estruturada (*Structured Query Language*)
TEEP	-	Efetividade Global (*Total Effectiveness Equipment Performance*)
UFPR	-	Universidade Federal do Paraná
VBA	-	Ferramenta do *Microsoft Officce* (*Visual Basic for Applications*)
xlsx	-	Extensão dos arquivos do *Software Microsoft Excel*

LISTA DE SÍMBOLOS

λ – autovalor de um vetor

Θ - ângulo entre duas retas

º - grau de classificação

Σ – matriz de covariância sigma

% - percentual

$\sum_0^i X$ – somatório

SUMÁRIO

1 INTRODUÇÃO

As organizações possuem como um dos seus principais objetivos a sua perpetuação no mercado, sendo indispensável para isso uma constante melhoria na qualidade de seus produtos e processos.

Segundo Nakajima (1988), uma afirmação mencionada nos ambientes industriais, é que a qualidade de um produto depende do seu processo de fabricação, e isso não deixa de ser verdade, porém na atual conjuntura tecnológica dos processos fabris, onde as linhas de produção são quase que completamente automatizadas, e as fábricas não tripuladas estão se tornando uma realidade, gradativamente, não apenas a qualidade, mas a produtividade, os custos, volumes de produção, entrega, saúde e segurança, dependem de equipamentos cada vez mais sofisticados. Porém, tal automatização dos equipamentos não visa eliminar a interação humana, pois ainda serão necessárias suas contribuições para as máquinas e para os processos, onde tais automações para serem bem-sucedidas requerem um sistema de manutenção confiável e apropriado, o que pode ser conquistado pela aplicação dos conceitos de Manutenção Produtiva Total (TPM).

Os objetivos gerais do TPM são os de reduzir os problemas dos equipamentos, buscando um aumento de sua confiabilidade de operação e por consequência conquistar índices satisfatórios de performance de máquina, qualidade de produtos, aumento de produtividade, redução de custos, entre outros. Para o monitoramento e análise de tais fatores, é utilizada uma ferramenta que permite o diagnóstico de um equipamento, tornando possível a análise das anomalias, o auxílio à tomada de decisões e o acompanhamento da evolução do processo por meio de indicadores, tal técnica é denominada de OEE (Eficiência Global do Equipamento). (NAKAJIMA, 1988).

Atualmente, com o crescimento das tecnologias digitais aplicadas no processo de fabricação, está ocorrendo uma mudança nas organizações em relação a sua eficiência de processos e de desempenho. Com o objetivo de integrar estes novos conceitos, as empresas estão adotando um novo modelo tecnológico de gestão denominado de Indústria 4.0, a qual é definida de forma genérica como sistemas de inteligência avançados para permitir o alto desempenho de produção, integrações rápidas e dinâmicas, e otimização dos tempos e tomadas de decisão. (SHIN, WOO e RACHURI, 2014).

As empresas estão enfrentando os desafios de lidar com os grandes volumes de dados gerados por esta nova era digital, os quais, por meio de técnicas analíticas inteligentes, podem propiciar uma melhor e mais rápida tomada de decisão para a melhoria de produtividade de uma organização. Porém, muitas empresas não estão preparadas para realizar a análise destes dados, ocasionando assim uma grande perda de informações para a manutenção da competitividade no negócio. (LEE, KAO e YANG, 2014).

A análise de dados com o objetivo de melhorar o desempenho das organizações está se tornando cada vez mais frequente e atraente para diversos modelos de negócio. Esse fomento é motivado por vários fatores, entre esses, a análise de grande quantidade de dados (*Big Data*) que está proporcionando novas visões e conclusões a respeito dos negócios. Tais fatores levam a um ambiente mais competitivo, onde pode ser conquistada uma melhoria de eficiência operacional, reduzindo custos, aumentando margens de lucratividade e perpetuando o negócio. (LIBES, SHIN e WOO, 2015).

A análise de dados, que possui como objetivo examinar os dados de forma bruta e lapidá-los para encontrar padrões e referências sobre determinadas informações, é um fator decisivo para o sucesso das organizações. (SHAO, SHIN e JAIN, 2015). Tais análises buscam a otimização dos diagnósticos de situações conhecidas e também de desconhecidas. Além disto, as análises preditivas necessitam de uma estrutura de informações para realizar a projeção de cenários futuros, por meio de uma grande quantidade de dados (Volume), com vários tipos de formatos de informações, estruturadas e não estruturadas (Variedade), além de aquisição e respostas rápidas aos dados para tomada de decisões em tempo hábil (Velocidade). (SHIN, WOO e RACHURI, 2014). De posse destas informações, as análises são realizadas tendo como princípio a confiabilidade dos dados (Veracidade), as quais retornarão informações valiosas para o negócio (Valor), permitindo uma tomada de decisão mais assertiva.

Nesse cenário, diversas técnicas são necessárias para lidar com essa quantidade de dados que vem crescendo a cada dia. A aplicação de técnicas estatísticas multivariadas, como Análise de Componentes Principais (ACP) em conjunto com Regressão Linear Múltipla, pode ser utilizada para a redução do número de variáveis preditoras da análise, tornando o modelo mais enxuto para a criação de uma equação de predição. A utilização de técnicas como as Redes Neurais Artificiais

(RNA), as quais possuem a habilidade de aprender sobre um processo por meio de dados de entrada de um ambiente, processando tal conhecimento e disponibilizando-o para um uso pretendido, torna-se um diferencial no processo de análise de dados, pois devido a sua elevada capacidade para resolução de problemas complexos, conseguem por meio do processamento de dados produzir saídas adequadas de informações, as quais em situações normais seriam incompreensíveis. (HAYKIN, 2001).

Nesse contexto, as abordagens multivariadas podem viabilizar excelentes resultados para análise de dados de equipamentos, neste novo cenário digital e repleto de informações disponíveis para análise, podendo gerar resultados confiáveis e de valor agregado para os processos de fabricação, além de possíveis previsões para decisões futuras mais assertivas, as quais visam gerar vantagem competitiva diferenciada e viável para o negócio.

Baseado nas circunstâncias expostas, a pergunta de pesquisa a ser respondida é: Como a abordagem multivariada pode ser empregada em modelos analíticos preditivos de OEE?

1.1 JUSTIFICATIVA

A magnitude de dados originados a partir das tecnologias aplicadas pelo conceito da indústria 4.0, geram uma potencial fonte de informações para o processo de fabricação. (SIVARAJAH, KAMAL, *et al.*, 2016). Atrelado a isso, tais dados armazenam considerável quantidade de conhecimento sobre o desempenho e eficiência dos equipamentos (OEE), os quais por sua vez possuem um potencial de exploração e análise, sendo valiosos para o entendimento do comportamento dos equipamentos, e importantes no processo de tomada de decisão em um ambiente fabril. (NAKAJIMA, 1988).

A análise de dados vem se tornando uma forte tendência que muitas organizações estão adotando com a finalidade de construir percepções valiosas a respeito do seu negócio e podendo assim realizar melhorias em seus processos, produtos e serviços. (LIBES, SHIN e WOO, 2015).

Com o passar do tempo tal conjuntura estará mais aplicada nas organizações, e este processo de análise de dados necessitará da implementação e utilização de ferramentas, como análises preditivas em abordagens multivariadas. (JUNQUÉ DE

FORTUNY, MARTENS e PROVOST, 2013). Assim, este trabalho tem o intuito de propor uma metodologia analítica preditiva empregada ao OEE, para que os produtos deste método possam agregar valor para tomadas de decisão assertivas, e assim proporcionar maiores vantagens competitivas no negócio de atuação.

1.2 OBJETIVOS

Para se alcançar resultados satisfatórios em relação aos propósitos da pesquisa, verificou-se como necessário a seleção de objetivos, os quais irão nortear o estudo e servir de base para a pesquisa. Tais objetivos são elencados abaixo:

1.2.1 Objetivo Geral

Propor um modelo para análise preditiva de OEE, por meio de uma abordagem multivariada aplicada em um banco de dados de registros de produção, em um processo de fabricação de embalagens de papel.

1.2.2 Objetivos Específicos

Os objetivos específicos do trabalho são:

a) Coletar, estruturar, organizar e analisar os dados originários de um processo de fabricação de embalagens de papel;
b) Aplicar técnicas estatísticas, como matriz de correlação e análise de regressão múltipla, para verificar a força de relação entre as variáveis;
c) Aplicar abordagens multivariadas, como análise componentes principais, análise de regressão múltipla e redes neurais artificiais, para elaboração de um modelo preditivo empregado em OEE;
d) Considerando uma família de produtos especifica, aplicar o modelo elaborado de predição do OEE, visando avaliar o seu comportamento e capacidade de predição.

1.3 LIMITAÇÕES DO TRABALHO

O estudo será desenvolvido com dados e informações oriundas de um banco de dados de produção de uma indústria de embalagens de papel e seu foco será na

aplicação da abordagem multivariada para a criação de um modelo analítico preditivo, aplicado no indicador de Eficiência Global do Equipamento (OEE).

1.4 CLASSIFICAÇÃO DA PESQUISA

Para Gil (2002), as pesquisas se iniciam por meio de algum problema ou indagação.

> Pode-se definir pesquisa como o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa é requerida quando não se dispõe de informações suficientes para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada. (GIL, 2002, p. 17)

Um projeto de pesquisa deve possuir uma disposição lógica e estruturada, as quais contemplem quais são os objetivos da pesquisa, a sua justificativa para ser realizada, qual método será empregado, além de como serão realizadas as coletas e análise dos dados, conforme demonstrado na FIGURA 1. (GIL, 2002).

FIGURA 1 – FLUXO DE RESOLUÇÃO DE PROBLEMA

FONTE: Martins (2012).

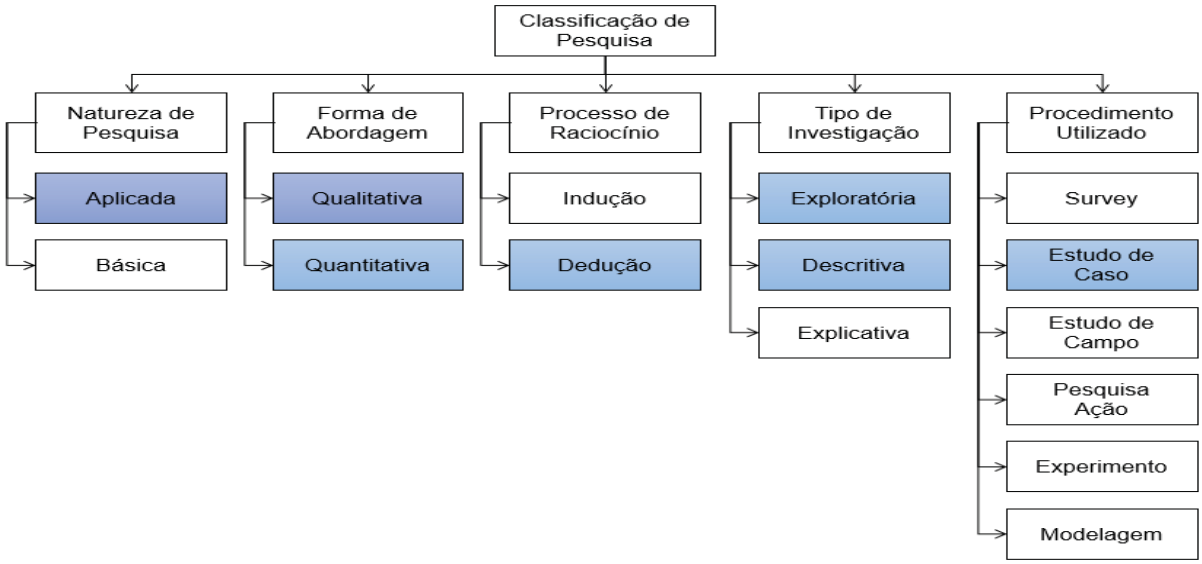
Segundo Gil (2002), as pesquisas devem ser classificadas, e está ação é realizada sob algumas premissas e baseadas nos objetivos gerais definidos. Essa classificação é composta por pesquisas do tipo exploratórias, descritivas e explicativas.

As pesquisas em engenharia de produção produzidas possuem grandes tendências às metodologias voltadas para Estudos de Caso (NAKANO, 2012). Também é fato que as abordagens mais utilizadas em engenharia de produção e em gestão de operações no Brasil são as que empregam metodologias de Estudo de Caso. (MIGUEL e SOUSA, 2012).

Conforme Gil (2002) o estudo de caso é uma forma de metodologia empírica, com o intuito de um estudo minucioso sobre uma determinada situação, onde se obtenha um diversificado e extenso conhecimento sobre o tema estudado.

Para a presente pesquisa será adotado o método de estudo de caso, por entender-se que este pode ser considerado como o mais adequado e vinculado ao objetivo geral da pesquisa. Na FIGURA 2, é demonstrada nas caixas destacadas a classificação da pesquisa.

FIGURA 2 – CLASSIFICAÇÃO DA PESQUISA

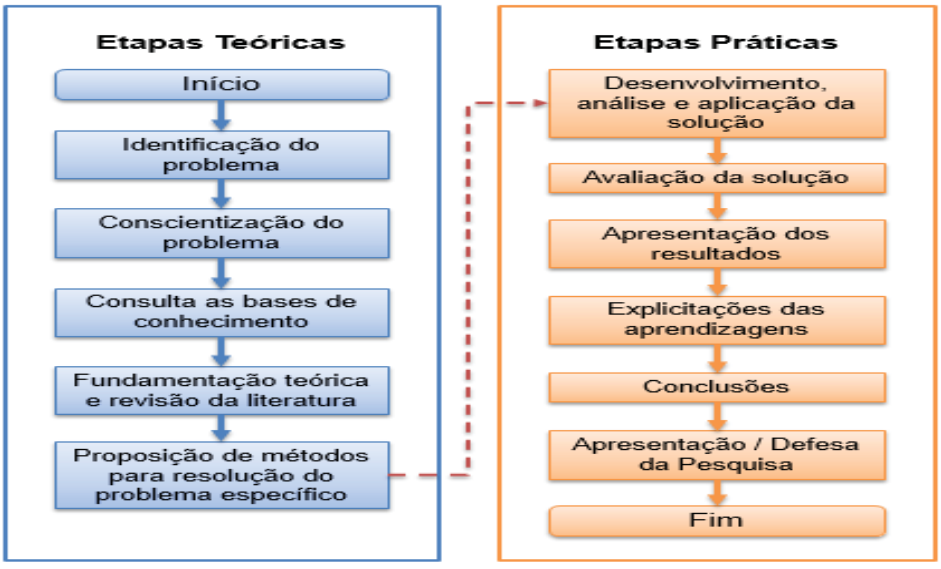


FONTE: Adaptado MIGUEL et al. (2012).

1.5 ETAPAS DA PESQUISA

O percurso metodológico de atividades e operacionalização dos objetivos para a execução da pesquisa, está proposto no fluxo demonstrado na FIGURA 3.

FIGURA 3 – ETAPAS DA PESQUISA



FONTE: O autor (2019).

1.6 ESTRUTURA DO TRABALHO

Buscando atingir os objetivos de pesquisa, o presente trabalho está dividido em capítulos para um melhor entendimento da metodologia de pesquisa aplicada.

- Capítulo 1: capítulo introdutório, o qual definiu-se o problema de pesquisa, a justificativa para o trabalho, o objetivo geral, os objetivos específicos e a classificação da pesquisa.
- Capítulo 2: nessa seção encontra-se a fundamentação teórica dos temas pesquisados e abordados no presente trabalho, sendo subdividido em conceitos de TPM, OEE, Indústria 4.0 e Análise Multivariada, com destaque para a análise de componentes principais e redes neurais artificiais.
- Capítulo 3: nesse capitulo são definidos os materiais e métodos que foram utilizados na pesquisa, como a forma que os dados foram coletados, a descrição das variáveis utilizadas nas análises, a forma de aplicação dos métodos multivariados e explanação da forma como foram conduzidas as análises.
- Capítulo 4: no quarto capítulo são apresentados os resultados e discussões obtidas com o processo de análise dos dados.
- Capítulo 5: no capítulo de conclusão comenta-se a sobre os resultados conquistados com o trabalho, explanando-se as possibilidades de estudos futuros.

2 REFERÊNCIAL TEÓRICO

No início dos trabalhos de pesquisa é necessária a realização de uma revisão bibliográfica, ou uma “varredura horizontal”, para conhecer quais as teorias existentes para o tema que será estudado, gerando assim conhecimento a respeito do assunto abordado. Na sequência é necessário um maior aprofundamento nos materiais encontrados, realizando nesta fase a chamada revisão de literatura, a qual propõe um alcance maior e mais profundo em conceitos já publicados e que possuem um maior grau de aderência ao objetivo da pesquisa, visando assim criar um embasamento para o trabalho. (FLEURY, 2012).

Neste trabalho, para a revisão da literatura foram utilizadas sequências lógicas de pesquisa, desdobramento e segregação de publicações voltadas ao tema de pesquisa. Em um primeiro momento realizou-se a revisão bibliométrica da literatura, em uma segunda etapa a revisão sistemática da literatura, as quais proporcionaram embasamento teórico sobre os temas pesquisados.

2.1 TPM

Após a segunda guerra mundial, a indústria japonesa baseou-se em conceitos e técnicas de manufatura originárias dos Estados Unidos, sendo que posteriormente o Japão chamou a atenção mundial pelo seu estilo de gestão. A indústria japonesa aprimorou os modelos americanos, tornou-se conhecida mundialmente pela qualidade superior de seus produtos e processos, e como consequência de tal feito exportou esses conceitos aprimorados para indústrias ocidentais. Umas das técnicas adotadas e aprimoradas pelos japoneses denomina-se Manutenção Produtiva Total, ou TPM. (NAKAJIMA, 1988).

TPM é originalmente uma ideia originada na Ford *Company*, mas que foi aperfeiçoada pelos japoneses na década de 1950 quando a manutenção preventiva foi introduzida no Japão. No ano de 1960, a Nippondenso, empresa parte do grupo Toyota, foi a primeira indústria japonesa a introduzir o conceito de manutenção preventiva em toda a sua planta. Neste conceito, os operadores focavam na produção e nas máquinas que compunham os processos de fabricação, e a equipe de manutenção era dedicada ao trabalho de manutenção dessas máquinas. No entanto, com o alto nível de automação dos equipamentos, a manutenção requeria cada vez

mais manutentores para dar apoio e suporte no processo, e isso começou a tornar-se um problema para a empresa, devido ao aumento de quadro necessário para a manutenção. Desta forma, a direção da empresa decidiu que as manutenções de rotina dos equipamentos seriam realizadas a partir daquele momento pelos próprios operadores, sendo esta técnica conhecida como manutenção autônoma, uma das características mais conhecidas e marcantes do TPM. (STAMATIS, 2010).

O TPM é uma metodologia de gestão e manutenção realizada diretamente pelos funcionários que atuam nos equipamentos e no local de trabalho, e possui como propósito maximizar a eficiência de um processo, mais especificamente de um equipamento. Em muitas ocasiões, essa interpretação gera um entendimento equivocado da gestão, pois o TPM é interpretado apenas como uma forma autônoma de manutenção no “chão de fábrica”, sem o envolvimento da alta gestão. Porém, para o TPM ser eficaz, este deve ser implementado com a participação efetiva da gestão, devendo ser apoiado de forma integral pelos membros da alta direção da organização. (NAKAJIMA, 1988).

Segundo Nakajima (1988), a aplicação da gestão da manutenção no processo de fabricação é focada nas atividades dos grupos autônomos, onde a palavra “total” no termo Manutenção Produtiva Total possui três significados que descrevem da melhor maneira a filosofia do TPM:

1. Participação total dos funcionários, ou envolvimento de uma forma completa dos funcionários que participam do processo de produção, sendo realizada pela manutenção autônoma e com participação dos operadores dos equipamentos, e também por meio de atividades pequenas melhorias desenvolvidas em grupos denominados de Kaizen.
2. Sistema de manutenção total, onde é inserido o conceito da manutenção preventiva, onde se estabelece um plano de manutenção para toda a vida útil do equipamento, e de melhorias nos equipamentos.
3. Total eficácia, o qual indica a busca da eficiência econômica ou da rentabilidade do TPM.

O TPM é um complemento fundamental para apoiar o processo de fabricação. Se o funcionamento de uma máquina não possui estabilidade, se o equipamento não possui confiabilidade de produção, e se a capacidade do processo não é sustentada,

a empresa precisa manter estoques extras para contrapor estas incertezas, e fluir sem ser interrompido. (STAMATIS, 2010).

Segundo Stamatis (2010), o conceito do TPM busca uma abordagem proativa, visando essencialmente evitar qualquer tipo de parada involuntária antes que a mesma possa ocorrer. Uma maneira de pensar em TPM é na prevenção da deterioração e na redução de manutenções corretivas nos equipamentos. Dentre os objetivos e as atividades do TPM, cinco possuem destaque:

1. Maximização da eficiência dos equipamentos;
2. Criação de um sistema de manutenção para prolongar a vida útil do equipamento;
3. Envolvimento de todos os departamentos nas fases de projeto, planejamento, uso ou manutenção de equipamentos;
4. Envolvimento ativo de todos os funcionários;
5. Promoção do TPM por meio de uma gestão motivacional.

Segundo Stamatis (2010), o TPM identifica os principais desperdícios do processo de produção, para depois trabalhar sistematicamente para eliminá-los, principalmente por abordagens de melhoria contínua. O TPM também possui oito pilares principais de atividades, sendo cada um desses configurado para atingir um alvo de “zero” desperdício. Esses pilares são:

1. Melhoria focada: destina-se a identificação e eliminação de desperdícios, sendo algum desses o transporte desnecessário, excesso de estoques, esperas, movimentações desnecessárias, excesso de produção, entre outros, os quais não agregam valor aos processos.
2. Manutenção autônoma: nesse pilar, o operador é a “peça-chave”. Isso envolve atividades de manutenção diárias realizadas nos equipamentos e que impedem a sua deterioração. Algumas atividades previstas neste pilar, são: limpeza, eliminação das fontes de sujeira, eliminação dos locais de difícil acesso, inspeção, lubrificação, padronização e gestão visual de anomalias.
3. Manutenção planejada: tem por objetivo atingir “zero” paradas involuntárias do equipamento.

4. Educação e treinamento: focado na preparação dos funcionários com o objetivo de aumentar a produtividade.
5. Gestão inicial de equipamentos/produto: possui o objetivo de reduzir o desperdício que ocorre durante o início da operação de uma nova máquina, ou a produção de um novo produto.
6. Manutenção da qualidade: implementa ferramentas da qualidade mais eficazes para o TPM, visando conseguir o “zero” defeito, tomando as medidas necessárias para prevenir perdas relacionadas ao produto e a qualidade dos mesmos.
7. Segurança, higiene e meio ambiente: visa atingir o “zero” acidentes relacionados ao trabalho, bem como as proteções necessárias para evitar impactos ambientais.
8. Escritório TPM: busca o envolvimento de todos os setores da organização nas práticas do TPM, pois os processos de escritório podem ser melhorados de forma semelhante aos ambientes fabris.

O propósito inicial do TPM, criado na década de 50, visa a melhoria da eficiência dos processos produtivos e dos equipamentos, porém naquela época tal conceito foi difundido e era aplicado em equipamentos com baixíssimas taxas de automação. Com o passar dos anos, as fábricas inteligentes e não tripuladas tornaram-se uma realidade. As produções de muitas empresas começaram a ser realizadas por robôs e produções automatizadas durantes as 24 horas diárias de trabalho. Anteriormente a isso, os requisitos de produção dependiam apenas dos processos, porém em um ambiente automatizado e repleto de robôs, a qualidade passa a depender das máquinas, como também a produtividade, custos, inventário, segurança, volume de produção, entre outros, depende muito dos equipamentos produtivos. (NAKAJIMA, 1988).

Segundo Nakajima (1988), uma maior automatização dos processos não eliminará a necessidade do trabalho humano, pois apenas as operações estão sendo automatizadas. Ainda, independente dos cenários e da evolução tecnológica, a manutenção e o processo de conservação dos equipamentos dependem fortemente das habilidades e da contribuição humana. A utilização de equipamentos tecnologicamente avançados requer competências mais especificas e técnicas do que as usuais, sendo assim necessário uma melhor gestão e organização da manutenção.

Essas práticas são previstas e praticadas na filosofia do TPM, o qual organiza os sistemas de gestão autônomo em equipes com operadores e manutentores dos equipamentos de produção, sendo que tal prática prevê e suporta os processos mais sofisticados tecnologicamente.

Nesse cenário que se aplica o TPM, também existe um processo de medição, o qual considera o desempenho do equipamento após observar a implementação de suas técnicas no processo, e se o equipamento adquire uma melhor eficiência. Para tal acompanhamento, recomenda-se a utilização do indicador de Eficiência Global do Equipamento, uma das métricas mais indicadas para essa prática, pois possui uma metodologia de melhoria contínua aplicada em seus conceitos. (STAMATIS, 2010).

2.2 OEE – EFICIÊNCIA GLOBAL DO EQUIPAMENTO

O conceito de Eficiência Global do Equipamento (*Overall Equipment Effectiveness* - OEE) foi introduzido por Seiichi Nakajima, por meio do livro Introdução ao TPM, publicado pelo instituto japonês de mantenimento de plantas (JIPM) no ano de 1984. (KENNEDY, 2018).

Um dos principais objetivos das atividades de melhoria da produção está relacionada ao aumento da produtividade, buscando a eliminação ou redução dos desperdícios, focando no valor agregado dos processos, aumento de volume de produção, melhorias da qualidade, redução de custos, atendimento dos prazos de entrega, entre outros fatores. Todos esses indicadores são altamente impactados pela eficiência dos equipamentos, sendo que anteriormente ao OEE eram observados de forma isolada e não poderiam trazer uma visão analítica e comparativa sobre a eficiência real do equipamento. (NAKAJIMA, 1988).

O OEE busca apoiar a produtividade de uma fábrica, demonstrando os principais problemas técnicos de um processo, além de todas as perdas que afetam de forma negativa o desempenho da produção, sendo o seu valor uma representação do potencial utilizado dos equipamentos. (MARKUS e STEINBECK, 2018).

O OEE é um conjunto de medidas que indicam a efetividade de uma operação de produção, sendo os resultados apresentados em um único número e de forma que possam ser comparados entre unidades fabris, departamentos, máquinas e processos. Sendo o OEE uma medida que identifica a eficiência e potencial de um

equipamento, demonstra e acompanha uma perda, e identifica oportunidades de melhoria. (STAMATISA, 2010).

Conforme Markus e Steinbeck (2018), um dos principais campos de aplicação do OEE é a priorização dos trabalhos de melhoria, focando as ações nas perdas de maior impacto para a empresa, e com potencial de aumento do OEE. Como principais objetivos do OEE podem ser citados:

- Aumento de produtividade;
- Redução de custos e aumento da rentabilidade;
- Redução das perdas e desperdícios;
- Aumento da vida útil do equipamento;
- Atingir, manter ou melhorar a vantagem competitiva da organização.

O OEE pode ser visto como um sistema para melhoria de gestão dos processos produtivos e dos equipamentos, não apenas como uma medida de comparação do desempenho dos equipamentos, além de ser uma ferramenta para o monitoramento e verificação da evolução da melhoria contínua nos equipamentos e processos. (KENNEDY, 2018).

Segundo Markus e Steinbeck (2018), o OEE pode ser calculado de duas maneiras distintas, porém retornando os mesmos resultados em ambos os casos. Na primeira forma é calculado pela razão entre o número de unidade produzidas num determinado período de tempo, em relação a quantidade potencial que poderia ser produzida no mesmo período, conforme a equação (1.1).

$$OEE = \frac{Número\ de\ Unidades\ Produzidas}{Número\ de\ Unidades\ Potencial\ de\ Produção} \times 100 \qquad (1.1)$$

Este modo de cálculo, apesar de ser mais simples, não demonstra os fatores que estão impactando na eficiência do equipamento, sendo mais utilizado para cálculos básicos e operacionais do OEE. Calculando o OEE desta forma mais simples, o resultado se torna mais estável e confiável contra tentativas de manipulação, e por consequência um melhor monitoramento, porém para uma análise mais estruturada esse contém poucas informações a respeito das perdas ocasionadas, sendo assim recomendável realizar o cálculo completo do OEE. (MARKUS e STEINBECK, 2018).

Segundo Stamatisa (2010), o OEE mede o quanto um equipamento é utilizado de forma eficaz, identificando restrições, e como estas se comportam e impactam no valor do OEE, sendo desdobrado a eficiência do equipamento em três componentes distintos, porém mensuráveis. Dessa forma, a eficiência do equipamento é medida pela multiplicação desses três fatores: disponibilidade de máquina, performance de produção e qualidade dos produtos fabricados. A equação (1.2) representa o cálculo do OEE.

$$OEE = Disponibilidade \times Performance \times Qualidade \quad (1.2)$$

2.2.1 Taxa de Disponibilidade

Um dos conceitos importantes para a disponibilidade de máquina é o tempo planejado de produção, sendo esse gerado pela diferença entre tempo disponível teórico e o tempo de paradas programadas, conforme demonstrado na equação (1.3). O tempo disponível teórico de produção é o tempo disponível total de uma operação ou também denominado de tempo solar, pois compreende as 24 horas de um dia, ou os sete dias da semana, ou as 52 semanas de um ano. Esse é o tempo que na teoria seria possível de produzir, porém desse são extraídas as denominadas paradas programadas, as quais não são contabilizadas para efeito do cálculo do OEE. Tais paradas programadas são aquelas previstas em calendário e acordadas previamente, podendo contemplar manutenções planejadas, falta de programação, finais de semana ou feriados, entre outras. Para o cálculo do OEE é utilizado o tempo planejado de produção, pois o OEE mede a eficiência de um equipamento em funcionamento, não quando o mesmo se encontra parado. (MARKUS e STEINBECK, 2018).

$$Tempo\ Planejado\ Produção = Tempo\ Teórico\ Produção - Paradas\ Programadas \quad (1.3)$$

A disponibilidade representa o quanto do tempo planejado de produção foi efetivamente utilizado para se produzir. O tempo efetivo de operação é o período em que realmente um equipamento ficou produzindo, não levando em consideração o tempo de inatividade do equipamento, no qual ocorreram paralisações não programadas, conforme a equação (1.4). (NAKAJIMA, 1988).

$$Tempo\ Efetivo\ Produção = Tempo\ Planejado\ Produção - Paradas\ não\ Programadas \quad (1.4)$$

A disponibilidade é uma porção do OEE que representa a razão entre as horas efetivas de produção pelas horas planejadas para a produção, sendo essa informação fornecida em forma de percentual, conforme demonstrado na equação (1.5). (NAKAJIMA, 1988).

$$Disponibilidade = \frac{Tempo\ Efetivo\ de\ Produção}{Tempo\ Planejado\ de\ Produção} \times 100 \quad (1.5)$$

A disponibilidade se concentra principalmente nas paradas involuntárias dos equipamentos, sendo muitas dessas relacionadas a manutenções, reparos, ajustes, entre outras. Por essa característica, a disponibilidade depende muito da confiabilidade e capacidade de manutenção do equipamento. Programas de manutenção planejada visam a maximização da disponibilidade de um equipamento pela redução de paradas involuntárias e aumento da confiabilidade, que gera como consequência menor tempo de inatividade e perdas de produção. Além das falhas e panes nos equipamentos, outro elemento que causa grandes perdas de disponibilidade são as trocas de operações, denominadas de *Setup*. (STAMATIS, 2010).

2.2.2 Taxa de Performance

Performance, ou desempenho, é a porção métrica do OEE que representa a razão entre a velocidade real praticada durante a produção em relação a velocidade teórica do produto em um determinado equipamento. Em outros termos, a performance representa um percentual da sua velocidade real em relação a velocidade teórica. (STAMATIS, 2010).

Segundo Nakajima (1988), o componente do OEE denominado de performance, refere-se a razão entre a velocidade de funcionamento real do equipamento em relação a velocidade ideal, conforme a definição do projeto do produto e de duas interações com o equipamento, conforme a equação (1.6).

$$Performance = \frac{Velocidade\ Real}{Velocidade\ Teórica} \times 100 \quad (1.6)$$

A taxa de desempenho do equipamento refere-se à quantidade máxima de produtos que podem ser produzidas em uma determinada quantidade de tempo,

levando-se em consideração o tempo de ciclo projetado de determinado produto. Em relação a isso, o mix de produção é um fator levado em consideração em relação as velocidades estipuladas, pois muitos fatores de projeto influenciam no rendimento potencial do equipamento, as interações entre a máquina e o produto podem levar a aumentos, ou reduções de velocidades, não existindo apenas uma velocidade padrão atrelada ao equipamento para qualquer tipo de produto. O melhor tempo de ciclo, e o tempo para cada tipo de produto normalmente não são idênticos, dependendo do projeto e vínculos do produto com a máquina para a definição da velocidade ideal, ou teórica. (MARKUS e STEINBECK, 2018).

Segundo Stamatis (2010), a performance refere-se à velocidade das máquinas, sendo um dos grandes objetos de estudo do OEE os motivos pelos quais os equipamentos não atingem a sua velocidade planejada, resultando em uma baixa taxa de desempenho, existindo duas principais formas de um equipamento não atingir um bom desempenho devido as perdas de velocidade:

1. Velocidade de funcionamento reduzida do equipamento, as quais ocorrem quando existem defeitos na máquina, fazendo com que o equipamento trabalhe em regime de velocidade menor do que o recomendado. A baixa velocidade também pode ser impactada pela mão-de-obra, a qual não possui segurança devido à baixa competência operacional, ou pela insegurança de geração do nível de qualidade inferior produzida no caso de velocidades mais elevadas. (STAMATIS, 2010).

   O desempenho e a velocidade máxima de operação possuem um limite superior máximo natural definido pelo projeto construtivo do produto, sendo nesses casos o nível máximo alcançável de desempenho de 100%, porém quando ocorrem determinações de velocidade teórica equivocadas e inferiores a velocidade real do equipamento, essas podem gerar uma taxa de performance maior do que 100%, fazendo com que o índice de OEE fique equivocadamente superior, sendo em até alguns casos maior que 100%. Para tanto é necessário uma revisão periódica e sistemática das velocidades teóricas por produto, para minimizar e evitar análises incorretas de OEE. (MARKUS e STEINBECK, 2018).

2. Pequenas paradas nos equipamentos, as quais se caracterizam por pequenas interrupções nos processos durante a produção, e que normalmente não são apontadas pela operação. Devido a esse fato são incluídas como perda de desempenho, pois impactam e reduzem a quantidade de saída do produto. Em casos de automação do equipamento, onde existam apontamento automáticos, essa perda pode ser contabilizada como disponibilidade, porém ainda impacta na performance dos equipamentos devido a geração das rampas de aceleração e desaceleração da máquina, as quais geram perdas de velocidade e consequentemente de desempenho. (STAMATIS, 2010). Em teoria as pequenas paradas correspondem na verdade a perdas de disponibilidade, pois não existe produção e a máquina fica parada durante esses períodos de tempo, porém devido a curta duração de segundos ou poucos minutos, na gama de um digito destas paradas e devido à dificuldade de registro das mesmas são geralmente registradas como perdas de velocidade. Muitas empresas adotam como pequenas paradas eventos de interrupções do equipamento menores que 5 minutos, por ser considerado um curto tempo de inatividade. (MARKUS e STEINBECK, 2018).

2.2.3 Taxa de Qualidade

Qualidade é a parte da métrica do OEE que representa a razão entre a quantidade de unidades produzidas com qualidade e as unidades totais produzidas. Em outras palavras, é o percentual resultante de peças que estão dentro das especificações definidas pelo cliente. (STAMATIS, 2010).

Segundo Nakajima (1988), o componente do OEE denominado de qualidade, refere-se a razão entre a quantidade de produtos com qualidade em relação a quantidade total produzida, conforme a equação (1.7).

$$Qualidade = \frac{Unidades\ Produzidas - Unidades\ Defeituosas}{Unidades\ Produzidas} \times 100 \qquad (1.7)$$

Como medida para o cálculo do OEE devem apenas serem considerados produtos bons, de qualidade. Se produtos defeituosos são produzidos, esses devem ser deduzidos do cálculo do OEE. Se a fábrica produz produtos fora do especificado, gera-se má qualidade, e neste caso é como se o equipamento estivesse parado, pois

não serão aproveitadas tais horas de produção. O OEE também é visto como uma medida de qualidade do processo, sendo que produtos defeituosos contam como erros do processo e sendo considerados como uma perda do OEE. (MARKUS e STEINBECK, 2018).

Problemas de qualidade podem acontecer a qualquer momento do processo de fabricação, desde que não sejam seguidas ou atendidas as especificações dos clientes e as condições de projeto ideais. Durante a fase de troca de pedido, inicialização ou ajustes do equipamento podem ocorrer desvios que irão gerar produtos defeituosos, denominados de refugo de s*etup*. (STAMATIS, 2010).

Embora no cálculo do OEE algumas medidas são relacionadas em unidade de tempo, a taxa de qualidade é medida por unidades de produção, sendo que a aplicação do cálculo dessa forma não gera nenhum inconveniente para o resultado final, pois as taxas do OEE de disponibilidade, performance e qualidade são expressas em fatores percentuais, neutralizando a influência da unidade de medida no cálculo. Porém, mesmo assim, é possível realizar a conversão de outras unidades de medida utilizadas em relação a horas de produção. (MARKUS e STEINBECK, 2018).

2.2.4 Maximizando a eficiência dos equipamentos

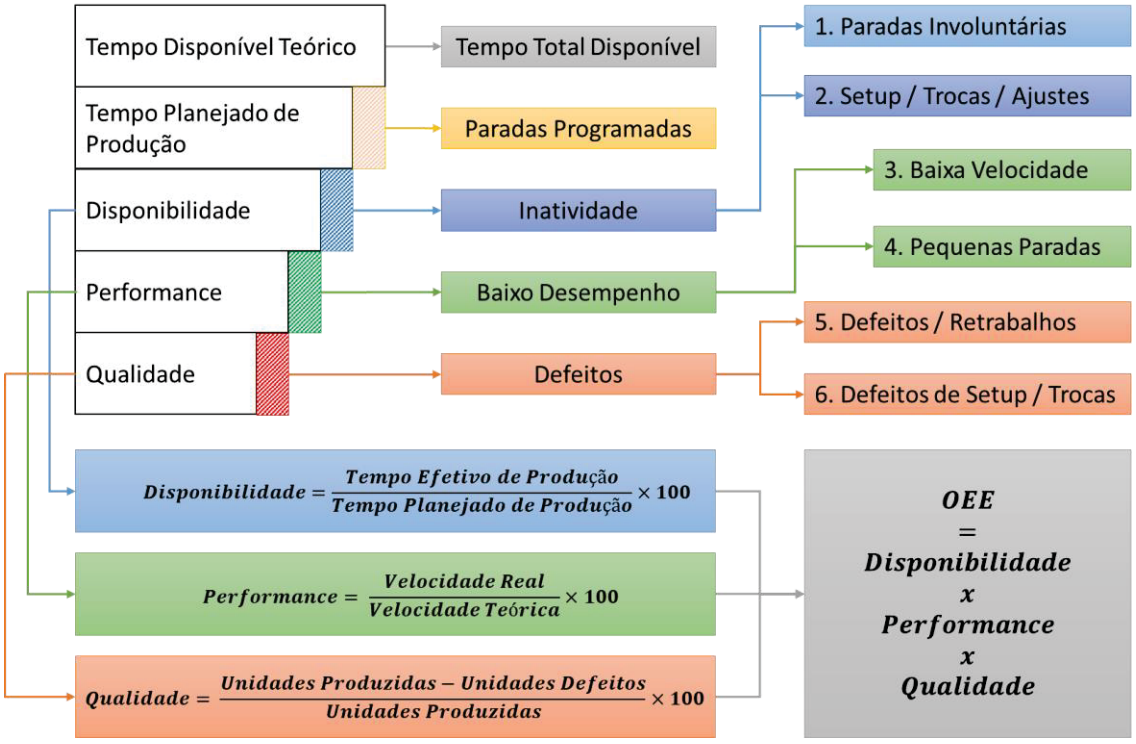
Segundo Nakajima (1988), visando alcançar bons índices de OEE, existe um esforço para eliminar os maiores obstáculos de perda de eficiência de um equipamento, denominados de as “seis grandes perdas” nos equipamentos. Essas são as principais responsáveis pelo baixo desempenho de uma máquina, e estão relacionadas a inatividade do equipamento (1 e 2), perdas de desempenho (3 e 4) e defeitos (5 e 6), sendo essas:

1. Falhas, panes, avarias ou paradas involuntárias do equipamento;
2. Trocas de pedido (*setup*) ou ajustes de máquina;
3. Velocidades reduzidas dos equipamentos, devido ao não atingimento da velocidade teórica ou ideal do equipamento;
4. Rampas de aceleração e desaceleração do equipamento, ocasionadas principalmente por pequenas paradas anormais do processo;

5. Defeitos do processo relacionados a problemas qualidade ou diferença entre o produzido e as especificações do cliente, ou retrabalhos;
6. Refugos gerados durante as trocas de pedido, ou durante o período de inicialização do equipamento.

Na FIGURA 4 é possível observar as seis grandes perdas do equipamento, bem como o conceito do OEE de forma sintetizada.

FIGURA 4 – CÁLCULO DO OEE E AS SEIS GRANDES PERDAS



FONTE: Adaptado de NAKAJIMA (1988).

Observado os conceitos aplicados ao OEE, inclusive a sua fórmula de cálculo, verifica-se que é praticamente impossível um OEE atingir uma eficiência de 100%, mesmo em um longo prazo, pois as falhas, defeitos, inatividades, trocas de pedido, baixas velocidades, entre outros, em algum momento farão com que sua eficiência fique abaixo da plenitude. Uma máquina ideal, ou com OEE de 100% seria um equipamento que nunca iria parar, ou não esperaria por nada, ou não geraria defeito algum, e também estaria sempre em velocidade máxima, desde de que houvesse demanda para esse período. Essa seria uma máquina com OEE pleno, porém tais

conjunturas são simplesmente suposições, pois tal condição é inexistente. (STAMATIS, 2010).

Em muitas fábricas o valor de referência do OEE é de 85%, e pode ser considerado como um excelente resultado de eficiência de um equipamento. Esse valor pode ser conquistado sendo necessárias algumas condições ideais, como uma taxa de disponibilidade superior a 90%, uma taxa de performance maior do que 95% e uma taxa de produtos de qualidade com mais de 99% de conformidade. (NAKAJIMA, 1988).

No entanto, independente do sistema de produção realizado na fábrica, alguns valores de OEE são referência para verificar o grau de evolução do processo. Caso um índice de OEE seja inferior a 30%, isso indica grande fraqueza do sistema de operação e baixa confiabilidade do sistema em atender os planos de produção. Valores intermediários de OEE, entre 30 a 50%, revelam um grau mediano de desempenho dos equipamentos. Valores muito baixos de OEE, próximo a 20%, porém com elevadas taxas de utilização, demonstram um sistema onde a produção não pode parar, porém não é demonstrada uma preocupação com as perdas ou desperdícios do processo. (MARKUS e STEINBECK, 2018).

A medição do OEE também é muito utilizada como indicador chave de desempenho de processos (KPI - *Key Performance Indicator*), sendo um acompanhamento e verificação sobre a evolução dos esforços de melhoria contínua em uma fábrica. O OEE pode ser utilizado como uma informação útil de referência, e além de ser utilizado para observação de eficiência, também pode fornecer elementos sobre a capacidade de um equipamento. (STAMATIS, 2010).

Um indicador muito utilizado em conjunto com o OEE para verificar a capacidade de um equipamento é o Desempenho Total do Equipamento Eficaz (TEEP - *Total Effectiveness Equipment Performance*) que representa o percentual do tempo total em que realmente um equipamento pode estar programado para a operação, ou seja, demonstra a capacidade de uma fábrica baseada no OEE. Para calculá-lo é necessário multiplicar o OEE pelo Tempo Disponível para Produção, assim é demonstrada a capacidade potencial não utilizada de um equipamento, sendo um incentivo para o aumento do OEE e por consequência o aumento de capacidade produtiva, porém tomando o devido cuidado para esse aumento de capacidade não se tornar um aumento desnecessário de produção, gerando superprodução e estoques. (STAMATIS, 2010).

Segundo Markus e Steinbeck (2018), um dos elementos críticos de sucesso do OEE é o processo de análise das perdas de desempenho dos equipamentos, para isso são recomendadas algumas etapas para uma melhor identificação das anomalias que prejudicaram o indicador, como:

- Identificar o valor absoluto do OEE;
- Considerar a evolução cronológica do OEE, verificando sua tendência;
- Verificar a variação geral do OEE;
- Estratificar o OEE em disponibilidade, Performance e Qualidade;
- Comparar cada elemento para verificar o maior contribuinte da perda.

Para as análises de tendência ou variabilidade do OEE, é importante entender os comportamentos cronológicos do processo, como a quantidade de horas trabalhadas, dias trabalhados dentro da semana ou do mês. Também pode ser levado em consideração para uma flutuação do OEE o mix e a sazonalidade dos processos, os quais afetam as interações entre máquina e produto. Além dessas situações, situações externas, como condições climáticas podem afetar alguns processos, como calor, umidade, verão, inverno, etc. Outras situações que podem estar atreladas ao comportamento humano também podem ter impacto no OEE, como dias da semana, horários de trabalho, turnos de produção, entre outros que podem afetar o comportamento da mão-de-obra. Essas flutuações podem demonstrar oportunidade para a melhoria dos processos e desperdícios que podem ser eliminados pela identificação e bloqueio das causas raízes e aplicação de contramedidas. (MARKUS e STEINBECK, 2018).

2.2.5 Coleta de Dados

Medir a eficiência global de um equipamento é um fator importante de monitoramento para ter a percepção de comportamento das máquinas, e quais as perdas que estão influenciado para a redução da eficácia da mesma. Para identificar anomalias e comportamentos anormais dos equipamentos é necessário medir e monitorar os equipamentos por meio de uma base de dados confiável, onde os problemas são identificados precocemente, e a eficácia das máquinas contabilizada. Com uma base de dados confiável, é possível medir e contabilizar o grau de evolução

das melhorias implementadas no processo. Para que essas avaliações sejam realizadas de forma segura é necessário que a coleta dos dados seja confiável, para se conquistar uma base segura e capaz de fornecer análise dos resultados de forma assertiva. (STAMATIS, 2010).

Para existir um sistema de melhoria rentável e um OEE de eficiência, é necessário que existam registros precisos da operação dos equipamentos, além de controles apropriados do processo. Em muitas empresas as paradas de máquina não são registradas de forma adequada. Em alguns processos de produção as paradas são apenas registradas quando extrapolam em muito tempo a indisponibilidade de um equipamento. Existe ainda empresas em que as pessoas que estão envolvidas nos processos entendem o apontamento de parada como uma burocracia e um desperdício de tempo. Para se possuir números confiáveis de disponibilidade, é essencial que os apontamentos de paradas e produção sejam precisos. Caso isso não ocorra, podem ser geradas interpretações equivocadas sobre o OEE. Em virtude disso, que os sistemas de apontamento de produção devem possuir a adesão dos funcionários, com características de serem simples e práticos. (NAKAJIMA, 1988).

A coleta de dados também pode ocorrer de forma automática. A coleta automática de dados é um termo utilizado para o monitoramento contínuo de dados reais a respeito de um processo ou negócio, sendo basicamente realizadas por sistemas informatizados, os quais são comumente ligados aos sistemas integrados de gestão (ERP - *Enterprise Resource Planning*) das empresas. Dessa forma, usando tecnologias, a base de dados também pode estar disponível de forma *on-line*, gerando assim informações e relatórios de OEE em tempo real. (MARKUS e STEINBECK, 2018).

Segundo Markus e Steinbeck (2018), a coleta automática de dados está muito atrelada a indústria 4.0, a qual está embasada nos avanços tecnológicos presentes nas empresas, que no caso do OEE mais especificamente, integrada as informações geradas e comunicadas pelos equipamentos, utilizando dispositivos que pelo conceito da indústria 4.0 darão significado diferenciado para o OEE, sendo possível eliminar obstáculos de informações que antes ficavam obscuras no ambiente fabril.

Uma das bases para a utilização do OEE na indústria 4.0, é a aplicação de sistemas de controles lógicos programáveis (CLP) atrelados a utilização de sensores para aquisição de dados e coleta de informações dos equipamentos, tecnologia de nuvem para processamento de dados com baixos custos de armazenamento,

terminais moveis para entrada e saída de informações em tempo real, *Big Data* e inteligência artificial para análise imediata dos dados. Com o conceito da internet das coisas (IoT), e tendo sensores conectados em uma rede, é possível o acesso, análise e gestão do OEE em tempo real, por meio de dispositivos como celulares, tabletes, entre outros conectados e carregando informações de uma nuvem de dados. Desta forma, não apenas o OEE de uma máquina pode ser observado, mas também desde todo o controle de uma fábrica, até a eficiência operacional de um produto pode ser medida e acompanhada em tempo real. (MARKUS e STEINBECK, 2018).

A análise de *Big Data* é um dos elementos da indústria 4.0 que possui grande utilização para a análise do OEE, devido as suas técnicas e métodos de análise. Utilizando técnicas de aprendizado de máquinas e algoritmos de redes neurais artificial, é possível reconhecer padrões a partir do banco de dados disponível pela coleta automática de dados, o que anteriormente era impraticável em virtude da grande quantidade de dados e variáveis disponíveis. Além disso, tais sistemas inteligentes possuem tecnologia para a realização de análises preditivas do OEE, ou seja, a previsão de possíveis situações futuras e tendências que possam impactar na eficiência dos equipamentos. Utilizando tais tecnologias, e existindo uma correta aplicação destes conceitos, é possível transmitir de forma quase que instantânea os índices, resultados, anomalias e dados para o OEE. Estas técnicas podem trazer novos conhecimentos para a gestão da fábrica, devido a padrões e exceções significativas que começam a serem identificadas, e que até então não eram do conhecimento das pessoas por não ser possível a visualização dos desvios e tendências de um processo. (MARKUS e STEINBECK, 2018).

2.3 A INDÚSTRIA DO FUTURO

A Indústria do Futuro tem trazido enormes benefícios para as organizações, onde novas tecnologias aplicadas as fábricas inteligentes, como a Internet das Coisas (IoT), os avanços da computação, a expansão das tecnologias sem fio, e as análises preditivas de grandes dados estão gerando acesso a informações que possuem potencial de melhorias para os processos de fabricação. (HE e WANG, 2017).

Embora existam diversos nomes usados para descrever a próxima geração de sistemas de fabricação, tal forma inteligente e avançada está sendo chamada de

Indústria 4.0. Para He e Wang (2017), sua essência é cada vez mais o emprego das tecnologias de informação e computação aplicadas ao ambiente fabril.

Para Sniderman, Matho e Cotteleer (2016), tal conceito ainda é recente e se originou da necessidade das empresas se tornarem competitivas, as quais almejavam melhorias em termos de qualidade e produtividade. Buscando a transformação e evolução, tanto no âmbito tecnológico, quanto no de processos, estas organizações se tornaram fábricas de ponta. No âmbito tecnológico, pode-se observar na FIGURA 5, de forma sucinta, as quatro revoluções industriais.

FIGURA 5 – GERAÇÕES INDUSTRIAIS

FONTE: GOTTWALD (2016).

Inicialmente, no fim do século 18 ocorreu a mecanização das fábricas por meio da utilização da máquina a vapor para aplicações industriais. No início do século 20, as fontes de energia migraram para a eletricidade, a qual trouxe como benefícios a produção em massa. A terceira revolução industrial ocorreu entre as décadas de 1970 e 2000, ocorreu o surgimento dos primeiros equipamentos e máquinas computadorizadas, sendo capazes de desenvolver atividades automáticas por meio de lógicas computacionais programadas, o qual só foi possível devido a aplicação de eletrônica e tecnologia da informação. A utilização de sistemas físicos-cibernéticos vem sendo aplicado em larga escala dentro das organizações, não apenas nos setores produtivos, mas também nas áreas de serviço. Tais eventos vêm

caracterizando esta fase como a quarta revolução industrial. (SNIDERMAN, MAHTO e COTTELEER, 2016).

As fábricas do futuro concentram seus esforços principalmente na otimização e inteligência de seus processos. Com esse intuito, um maior discernimento pode ser conseguido pela integração de sistemas que tem um impacto direto no desempenho dos equipamentos. Essa interação está relacionada a um processo contínuo de autoconhecimento e autoaprendizagem das máquinas, e consequentemente a uma melhoria no desempenho dos processos. (LEE, KAO e YANG, 2014)

Um dos grandes focos das fábricas inteligentes é possuir dados a respeito dos seus processos, e se possível em tempo real. O objetivo desta nova forma de pensamento é que tais informações possam contribuir de forma oportuna e precisa para a tomada de decisão dentro das organizações. Em virtude disso, a análise de grandes quantidades de dados (*Big Data Analytics*) visa a contribuir significativamente para o avanço da indústria do futuro. (HE e WANG, 2017).

A recente criação e utilização do conceito da internet das coisas, atrelado aos sistemas cyber-físicos (CPS), onde a utilização de sensores que coletam informações a respeito do processo, tem gerado um ambiente com uma vasta quantidade de dados na indústria. (LEE, KAO e YANG, 2014). O conceito de IoT prevê a conexões entre diversos objetos, equipamentos, ferramentas e ambientes em uma organização, os quais ocorrem por meio de sensores inteligentes que conectam e integram o mundo físico ao cibernético. Todo esse processo resulta em uma rede conectada, que permite a transmissão de dados e informações, de modo fluido e imediato, de máquina para máquina, ou de máquina para usuários. (SNIDERMAN, MAHTO e COTTELEER, 2016).

O uso de técnicas de mineração de dados traz informações relevantes por meio dos dados históricos colhidos e armazenados. Agregando a computação em nuvem a todo esse processo e com análises avançadas destes dados, as indústrias do futuro serão capazes de alcançar um sistema de informação em toda sua cadeia de produção, o que servirá de base para análise, previsões futuras e de tomada de decisões, criando um valor sustentável dentro de uma indústria 4.0. (LEE, KAO e YANG, 2014).

A aplicação dos conceitos da internet das coisas em um negócio traz como resultado um grandioso e complexo número de dados condensados e inseridos em um sistema de dados, denominado de *Big Data*. (YIN e KAYNAK, 2015).

A coleta, distribuição e gestão dos dados em um ambiente de *Big Data* são fundamentais para alcançar os *status* de máquinas autônomas ou inteligentes. Para isso se reforça a importância de se possuir tais dados de forma acessível para que os mesmos possam ser utilizados e analisados, reforçando a importância de aproveitar a flexibilidade adicional e capacidades ofertadas pela computação em nuvem, onde sua utilização nestes ambientes é inevitável. (LEE, KAO e YANG, 2014).

Toda a massa de dados pode estar disponível na chamada “nuvem”, a qual é definida como uma arquitetura de distribuição de informações e que possui como função disponibilizar de forma segura, rápida e conveniente, o armazenamento de dados. (HASHIZUME, et al., 2013).

A computação em nuvem é uma tecnologia que está em rápido e constante crescimento, e que se estabeleceu nas organizações, principalmente nas indústrias de tecnologia da informação e de negócios. Para Hashem et al. (2015), a computação em nuvem é um modelo que permite a onipresença das informações, e com acessos sob demanda, atrelado a recursos de computação, como redes, servidores, aplicações, serviços, entre outros, que podem ser rapidamente acessados e fornecidos, com o mínimo de esforço para gerenciar os dados contidos em seu provedor. (HASHEM, et al., 2015).

A computação em nuvem e *Big Data* são conjugadas, pois a computação em nuvem fornece ao analista de dados a possibilidade de consultar os dados necessários para a realização das suas análises. Após, retorna as informações resultantes do processo de forma antecipada para a tomada de decisão. (HASHEM, et al., 2015).

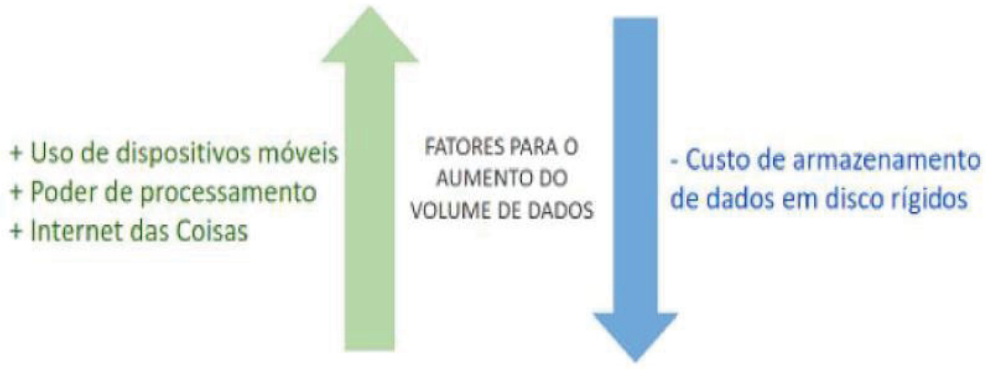
Segundo Talia (2013), a computação em nuvem é uma das formas mais adequadas de plataforma para armazenamento e análise de *Big Data*.

O desenvolvimento da indústria 4.0 tem sido acompanhado de descoberta de novas tecnologias para apoiar o processo de produção, com o propósito de serem ágeis, dinâmicas e adaptáveis as atuais demandas e exigências do mercado. Estas exigências dos consumidores estão atreladas a inovação, qualidade e variedade dos produtos, além da velocidade de entrega. Tais requisitos forçam as fábricas a terem capacidade de autoconsciência, auto previsão, auto comparação, auto reconfiguração, e automanutenção. Acompanhando essas necessidades de produção, tecnologias em desenvolvimento estão em foco e ganhando demasiada atenção nos meios acadêmicos e industriais. (LEE, KAO e YANG, 2014).

2.3.1 *Big Data*

Até o início do século XXI, grande parte dos dados armazenados eram em formato analógico, como por exemplo as fitas cassete e os disquetes, os quais eram frágeis, além de dificultar o armazenamento, proteção, distribuição, entre outros recursos que estão disponíveis atualmente nos arquivos digitais. Nos dias atuais é fácil de perceber toda essa transformação ocorrida com os dados e quanto a sua administração evoluiu, basta observar, por exemplo, os recursos disponíveis para compartilhamento de fotos, vídeos, documentos, entre outros. O advento da internet foi um dos percursores de toda essa evolução, porém sem as novas tecnologias que permitiram o aumento do armazenamento e manipulação dos dados, atrelados a redução dos custos para isso, essa evolução não seria possível, conforme demonstrado na FIGURA 6. (MARQUESONE, 2016).

FIGURA 6 – PRINCIPAIS FATORES PARA O AUMENTO DO VOLUME DE DADOS



FONTE: MARQUESONE (2016).

O termo *Big Data* é relativamente novo nas organizações, sendo utilizado para se referir ao aumento do volume de dados que são difíceis de armazenar, processar e analisar por meio de tecnologias de banco de dados tradicionais. (HASHEM, et al., 2015).

Segundo Gartner (2018), *Big Data* são dados que possuem alto volume, com alta velocidade e alta variedade de informações, os quais exigem inovação e economia para o seu processamento, permitindo desta forma um ambiente com uma melhor compreensão para a tomada de decisão de um processo.

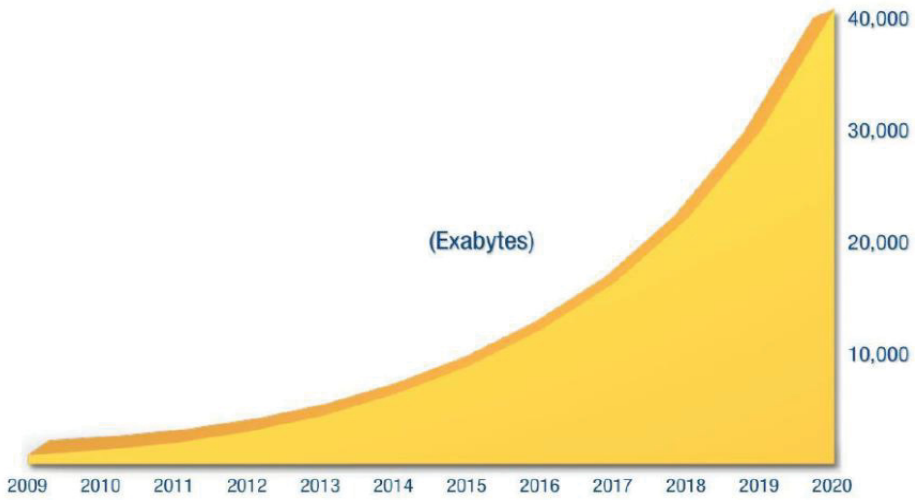
Segundo Hashem et al. (2015), sob o ponto de vista analítico, o *Big Data* é um conjunto de técnicas e tecnologias que exigem novas formas de integração para

descobrir grandes valores ocultos de grandes conjuntos de dados que são diversos, complexos e de grande escala.

Big Data é uma coleção de grandes e variados conjuntos de dados, com uma imensa diversidade de tipos de informações, o que se torna difícil para processar usando abordagens de processamento de dados tradicionais. (PHILIP CHEN e ZHANG, 2014).

Segundo Olshannikova et al. (2015), as fontes de dados estão presentes em todo o mundo e são geradas por diversas formas, como identificadores de rádio frequência (RFID), instrumentos de medição, mídias sociais, mensagem, sensoriamento remoto, dados de localização, dados de dispositivos móveis, gravações de áudio e vídeo, entre inúmeros outros. Em virtude do advento e utilização em massa do conceito da Internet das Coisas, estima-se que nos próximos anos esses dados irão crescer de forma exponencial. (HE e WANG, 2017).

FIGURA 7 – PREVISÃO EXPONENCIAL DE CRESCIMENTO DE DADOS DE 2009 A 2020



FONTE: IDC (2012).

Segundo IDC (2014), e conforme demonstrado pela FIGURA 7, a cada dia o mundo produz cerca de 2,5 quintilhões de *bytes* de dados, onde 1 quintilhão de *bytes* equivale a 1 bilhão de *gigabytes*. Segundo a projeção, no ano de 2020, mais de 44 *zettabytes*, ou 44 trilhões de *gigabytes*, de dados terão sido gerados. A TABELA 1 demonstra a relação entre as principais unidades de medidas computacionais utilizadas e seu equivalente em *bytes*.

TABELA 1 – UNIDADES DE MEDIDA PARA ARMAZENAMENTO DE DADOS

Nome	Símbolo	Equivalente em *Bytes*
Byte	B	10^0
Kilobyte	kB	10^3
Megabyte	MB	10^6
Gibabyte	GB	10^9
Terabyte	TB	10^{12}
Petabyte	PB	10^{15}
Exabyte	EB	10^{18}
Zettabyte	ZB	10^{21}
Yottabyte	YB	10^{24}

FONTE: Adaptado IBM (2018).

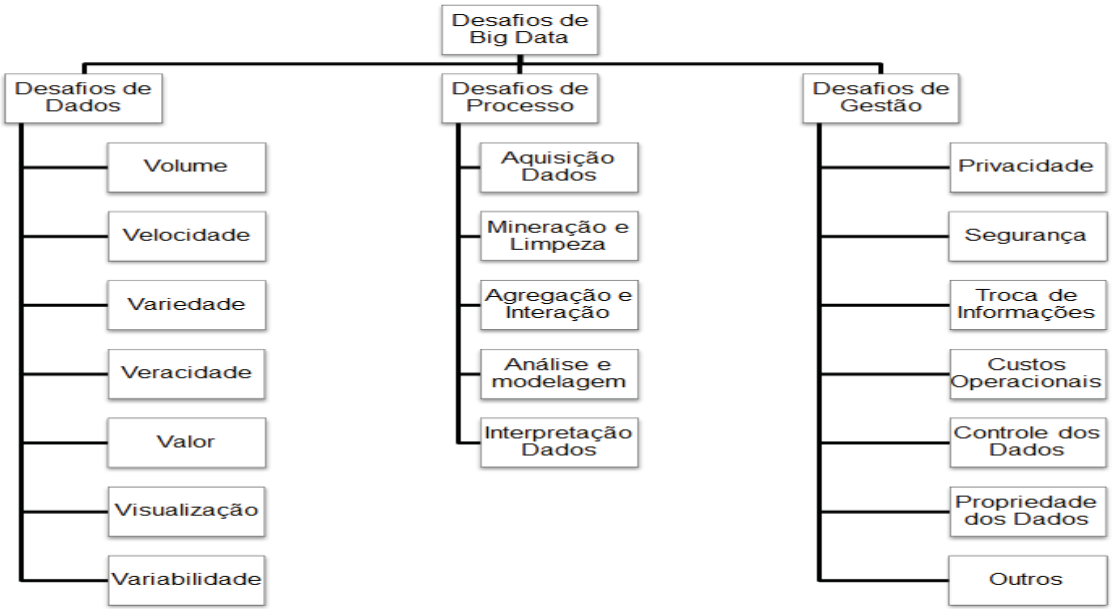
2.3.1.1 Desafio de *Big Data*

Segundo Philip Chen e Zhang (2014), *Big Data* traz muitas oportunidades atraentes, mas junto com estas são trazidos desafios para serem enfrentados. Segundo Sivarajah et al. (2016), grande parte desses desafios são características do próprio *Big Data*. Philip Chen e Zhang (2014), afirmam que se esses desafios não puderem ser enfrentados, estarão sendo desperdiçadas grandes e potenciais oportunidades. Tais dificuldades referem-se à busca, captura, armazenamento, compartilhamento, análise e visualização de dados.

Segundo o MIT Technology Review (2013), o conceito de grande quantidade de dados é complexo, uma vez que uma serie de dados que parece ser gigantesca hoje, muito provável que seja considerada pequena em um futuro próximo. O desempenho dos computadores e a velocidade dos processadores estão aumentando consideravelmente, porém não no mesmo ritmo. Os métodos de processamento de informação e os modelos de análise estão em um desenvolvimento mais lento, sendo que muitas técnicas e tecnologias de *Big Data* não são tão eficazes, especialmente no processamento em tempo real. (PHILIP CHEN e ZHANG, 2014).

Segundo Zicari (2014), os grandes desafios da *Big Data* podem ser agrupados em três categorias principais, conforme apresentado na FIGURA 8.

FIGURA 8 – CLASSIFICAÇÃO DOS DESAFIOS DE *BIG DATA*



FONTE: Adaptado de SIVARAJAH et al. (2016).

- Desafio de dados: referem-se às características dos dados em si.
- Desafios de processo: relacionados a uma série de técnicas para a captura, integração, transformação dos dados, além da seleção dos modelos mais adequados para análise dos dados, com posterior fornecimento de resultados.
- Desafios de gestão: relacionados à privacidade, segurança, governança e aspectos éticos dos dados.

2.3.1.2 Desafio de dados

Desafios de dados são os grupos de desafios relacionados com as características dos próprios dados, formando assim o conceito dos 7V's de *Big Data*. (SIVARAJAH, et al., 2016).

1. Volume: o volume se refere a grande quantidade de dados gerados principalmente por máquinas e dispositivos eletrônicos. Tal característica determina o tamanho de um conjunto de dados. (OLSHANNIKOVA, OMETOV, *et al.*, 2015).

O termo volume está relacionado às grandes quantidades de dados representadas em *terabytes, petabytes, zettabytes* ou ainda mais. Essa grande quantidade de dados é um grande desafio pela sua própria essência. (SIVARAJAH, KAMAL, *et al.*, 2016).

O que pode ser considerado como *Big Data* atualmente pode perder ficar desatualizado em futuro breve, pois a capacidade de armazenamento de dados tende a aumentar nos próximos anos, fazendo com que as quantidades de dados existentes se tornem ínfimas se comparadas com a de um cenário futuro. (GANDOMI e HAIDER, 2014).

2. Variedade: segundo Gandomi e Hader (2014) variedade refere-se à heterogeneidade estrutural em um conjunto de dados. Para Sivarajah (2016), a variedade se refere a múltiplos formatos de dados, como por exemplo, textos, imagens, áudios, vídeos, entre outros. Devido ao elevado volume de dados e de suas origens, as informações não possuem um padrão ou formato específico, sendo capturados de diversas formas e de diferentes fontes.

   Para Olshannikova et al. (2015), a variedade dos dados também está relacionada ao formato do conteúdo, sendo estruturado, ou não estruturado.

   Em um processo de fabricação são gerados inúmeros dados, de diferentes etapas de operação e de diferentes dispositivos. Tais dados muitas vezes são gerados por inúmeras fontes, com características diferentes, unidades de medidas variadas e frequência de coleta diferenciada. (HE e WANG, 2017).

3. Veracidade: segundo Sivarajah et al. (2016), a veracidade está ligada as imprecisões e inconsistências de um conjunto de dados. Para Olshannikova et al. (2015), a veracidade está ligada a complexidade dos dados que pode levar a uma falta de precisão e qualidade das informações.

   Veracidade não trata apenas a qualidade dos dados, mas também a sua compreensão, devido às discrepâncias em suas informações. (ZICARI, 2014). Desta forma, a necessidade de lidar com tais dados torna essa realidade mais um desafio de *Big Data*, pois durante a análise de dados,

existirão dúvidas, distorções, imprecisões, incertezas, dados falsos, confusões e ausência de informações. (GANDOMI e HAIDER, 2014).

Segundo Shao, et al. (2015), durante a análise de dados, existe uma grande quantidade de ruídos misturados com informações úteis. O desafio muitas vezes enfrentado é de decidir qual informação utilizar, separando a "sujeira" das demais informações.

Nos processos de produção, a veracidade está ligada a qualidade dos dados, geradas principalmente por ruídos, falta de dados, atrasos de informações oriundas de sensores (IoT) ou dispositivos, falta de sincronia dos dados entre dispositivos e *softwares*, entre outras. (HE e WANG, 2017).

4. Valor: segundo Sivarajah et al. (2016), valor se refere ao conhecimento extraído a partir de grande quantidade de dados, sem a perda de informações para o usuário final. Essa é uma característica essencial e mais importante de *Big Data*, pois dentro de uma magnitude de dados, existem informações valiosas, e extraí-las é um processo muito precioso.

5. Velocidade: a velocidade está relacionada ao quão rápido os dados são gerados, capturados e transmitidos. (SHAO, SHIN e JAIN, 2015).

   Segundo He e Wang (2017), além da taxa pela qual os dados são criados, a velocidade também se refere à quão rápido são analisados e colocados em prática.

   O aumento de dispositivos digitais elevou a taxa de criação de dados e por consequência velocidade de processamento dos mesmos, gerando assim uma necessidade cada vez maior de análises mais rápidas e simultâneas. (GANDOMI e HAIDER, 2014).

6. Visualização: a visualização está relacionada em como os dados, informações e conhecimentos são apresentados para os usuários. Busca-se com a visualização formas de demonstrar de maneira mais eficaz e intuitiva os valores extraídos de uma análise de dados. (SIVARAJAH, KAMAL, *et al.*, 2016).

   Segundo Philip Chen e Zhang (2014), a visualização tem por objetivo transmitir as informações de forma mais fácil, por meio de diferentes tipos

de gráficos, fornecendo conhecimentos ocultos e escondidos nas complexas estruturas de dados analisadas.

7. Variabilidade: a variabilidade é muito confundida com a variedade. Tem relação com o contexto em que o dado está inserido, com as mudanças de opiniões e de cenários. Toda vez que um dado é inserido em uma análise, porém oferece um significado diferente a cada interação, essa é a variabilidade. Os dados gerados e coletado por humanos constituem o maior valor de variabilidade, pois possuem altas taxas de alteração, além de também estar relacionada às análises de sentimentos, que no caso, por exemplo, dependendo do momento, uma palavra pode possuir um contexto totalmente diferente dentro de um determinado cenário, mas possuir total lógica em outro. (SIVARAJAH, KAMAL, *et al.*, 2016).

2.3.1.3 Desafios de processo

A segunda vertente de desafios de *Big Data* está relacionada ao processamento dos dados, focando em todos os processos desde a geração e captura dos dados, até a apresentação dos resultados finais. (OLSHANNIKOVA, OMETOV, *et al.*, 2015).

Como a maioria dos dados contidos normalmente não são estruturados, o processamento de tais informações proporciona um grande e complexo desafio em ralação a análise dos dados. (SIVARAJAH, KAMAL, *et al.*, 2016).

Segundo Sivarajah et al. (2016), os desafios de processo podem ser classificados em 5 etapas:

1. Aquisição dos dados: focado na obtenção e armazenamento dos dados oriundos de fontes confiáveis;
2. Mineração e limpeza dos dados: aplica-se a extração e limpeza dos dados, principalmente aos com características de não estruturados;
3. Agregação dos dados: compreende no agrupamento e integração dos dados extraídos e limpos;
4. Análise e modelagem dos dados: os dados estão aptos para serem analisados e incrementados em um modelo matemático*;*

5. Interpretação dos dados: apresentação dos dados após as análises, a fim de adquirir valor e conhecimento das informações obtidas.

2.3.1.4 Desafios de gestão

Segundo Olshannikova et. al. (2015), o último tipo de desafio está relacionado com gestão de dados. Estes desafios normalmente se referem às formas de gestão dos dados, onde em sua maioria possuem regras, leis e políticas previstas em âmbitos nacionais e internacionais, regendo o uso de tais informações. Para Sivarajah et al. (2016), essa gestão de dados refere-se a como acessar, gerenciar e governar dados.

Existem vários desafios de relacionados à gestão dos dados, dos quais podem ser mencionados: privacidade, segurança, governança, transação de informações, custos operacionais, controle de dados, entre outros. (SIVARAJAH, KAMAL, *et al.*, 2016).

2.3.1.5 Análises de *Big Data* (*Big Data Analytics* – BDA)

Segundo Hashem et. al. (2015), transformar grandes quantidades de dados em um modelo adequado para a análise é um dos grandes desafios de *Big Data*.

A análise de dados pode ser vista como a ciência de examinar os dados para descobrir padrões ocultos, correlações desconhecidas e outras informações úteis que podem ser utilizadas para tomar melhores decisões ou desenvolver soluções eficazes para o negócio de atuação. Dentro do âmbito das fábricas inteligentes, a análise de *Big Data* é um dos conceitos mais importantes, pois pode se consolidar como o alicerce da inovação, do aumento de produtividade e competitividade, pois é fonte de informações para a tomada de decisões. (HE e WANG, 2017).

Por meio da análise de *Big Data*, utilizando como base os históricos de produção, as situações atuais e das projeções futuras, pode auxiliar na melhoria de toda a cadeia de produção mediante elucidações sobre padrões, tendências, ineficiências e riscos potenciais para o negócio. (SHAO, SHIN e JAIN, 2015).

Para capturar valor e informações relevantes por meio de *Big Data* é necessário desenvolver tecnologias, modelos e técnicas para analisá-lo, sendo necessário considerar diversas disciplinas do conhecimento, como ciências da computação, matemática, estatística, entre outras especialidades relacionadas ao foco da análise. (PHILIP CHEN e ZHANG, 2014).

2.3.1.6 Métodos analíticos de *Big Data*

Big Data possui um imenso conjunto de dados brutos, mas que não oferece muito valor se não for processado. Para conquistar tais informações, as organizações necessitam utilizar técnicas adequadas e métodos eficientes para lapidar os complexos volumes de dados brutos existentes e transformá-los em informações transparentes e quantificáveis que auxiliarão e darão sentido ao processo de tomada de decisão e de desempenho organizacional, além de descobrir novas inconsistências e oportunidade. (SIVARAJAH, KAMAL, *et al.*, 2016).

Segundo Gartner (2014), existem três principais métodos de análise de dados: a análise descritiva, análise preditiva e análise prescritiva.

Na FIGURA 9 são relacionados os principais modelos aplicados em *Big Data*. Da esquerda para a direita indicam um progresso de utilização dos modelos de análise de dados para conquistar os níveis mais elevados e mais valiosos de apoio à decisão.

FIGURA 9 – FLUXO DE ANÁLISE EM *BIG DATA*

FONTE: Adaptado DEZYRE (2016).

2.3.1.6.1 Análise descritiva

A análise descritiva é o método de identificar o que aconteceu, ou o que está acontecendo, proporcionando de forma resumida visões diferenciadas do processo, além de encontrar padrões de tendências no mesmo. (SHAO, SHIN e JAIN, 2015).

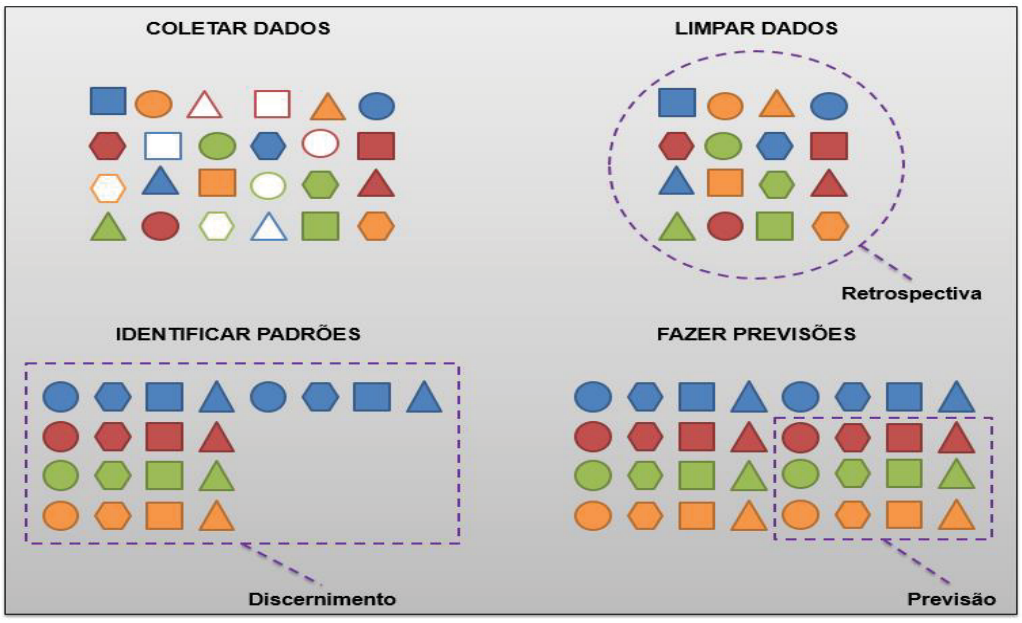
Para Sivarajah et. al. (2016), tal método faz a análise dos dados históricos, contendo o resumo e a descrição dos padrões de conhecimento, trazendo por meio de relatórios, a situação atual do processo, de forma que padrões e exceções se tornam evidentes.

Este é o modelo de análise mais simples, pois utiliza abordagens estatísticas simples, como média, desvio padrão, variância, entre outros. As análises descritivas vêm sendo utilizadas há algum tempo pelas empresas por meio de relatórios gerenciais, painéis de gestão a vista, entre outros, sendo aplicadas como ferramentas de acompanhamento do sistema de gestão. (SIVARAJAH, KAMAL, *et al.*, 2016).

2.3.1.6.2 Análise preditiva

A análise preditiva busca a determinação de visões e possibilidades futuras, realizando previsões e modelagens estatísticas, baseadas nos dados atuais e históricos das organizações, conforme exemplo demonstrado na FIGURA 10. (SIVARAJAH, KAMAL, *et al.*, 2016).

FIGURA 10 – ANÁLISE PREDITIVA



FONTE: Adaptado GAMESAUCE (2017).

Segundo Junqué de Fortuny et. al. (2013), com o passar do tempo vão sendo criadas novas formas de análise de *Big Data* com o intuito de buscar tendências e informações a partir de uma grande quantidade de dados. As organizações buscam tirar proveito destas informações, para se tornarem mais competitivas perante as demais, as quais não possuem acesso a tais informações, ou não tem habilidade de como fazê-lo. Com este intuito, a análise preditiva é uma das melhores formas para fornecer dados e informações para a tomada de decisão.

A análise preditiva é baseada em um conjunto de dados para os quais se pretende buscar um valor, porém não é simples de atingir altos níveis de precisão e desempenho preditivo sem um grande número de dados. A previsão destes modelos melhora quando se aumentam o número de situações observadas, esse é um diferencial competitivo para as organizações que possuem grande quantidade de dados e sabem como analisá-los. (JUNQUÉ DE FORTUNY, MARTENS e PROVOST, 2013).

Junqué de Fortuny et. al. (2013) afirmam que aumentar a variedade de dados é uma forma de aumentar a quantidade de dados coletados, proporcionando assim modelos mais ricos e com melhores capacidades preditivas. Porém, devido a fatores comportamentais, nos dados extraídos de ações humanas as taxas de sinal-ruído são muito altas, podendo gerar alta variabilidade dos dados.

É necessário tomar cuidado com os dados coletados para a análise, pois maiores quantidades de características podem levar a uma maior variação na modelagem preditiva, podendo gerar erros de predição. (JUNQUÉ DE FORTUNY, MARTENS e PROVOST, 2013).

2.3.1.6.3 Análise prescritiva

A análise prescritiva se concentra em como as ações previstas podem ser realizadas e quais as suas consequências para o negócio, além de identificar as políticas e recursos necessários para atingir os resultados desejados. (SHAO, SHIN e JAIN, 2015).

Este modelo busca determinar as relações de causa e efeito entre as políticas aplicadas na organização e os resultados almejados, por meio da otimização dos seus modelos de processo e negócio, agindo com base nas informações fornecidas pela análise descritiva e pelas previsões dos modelos preditivos. A análise prescritiva

contribui para a melhoria contínua dos processos e modelos de negócio, auxiliando no processo de tomada de decisão, criando contramedidas e avaliando o seu impacto em relação às estratégias das organizações. (SIVARAJAH, KAMAL, *et al.*, 2016).

2.3.1.7 Aplicações estatísticas

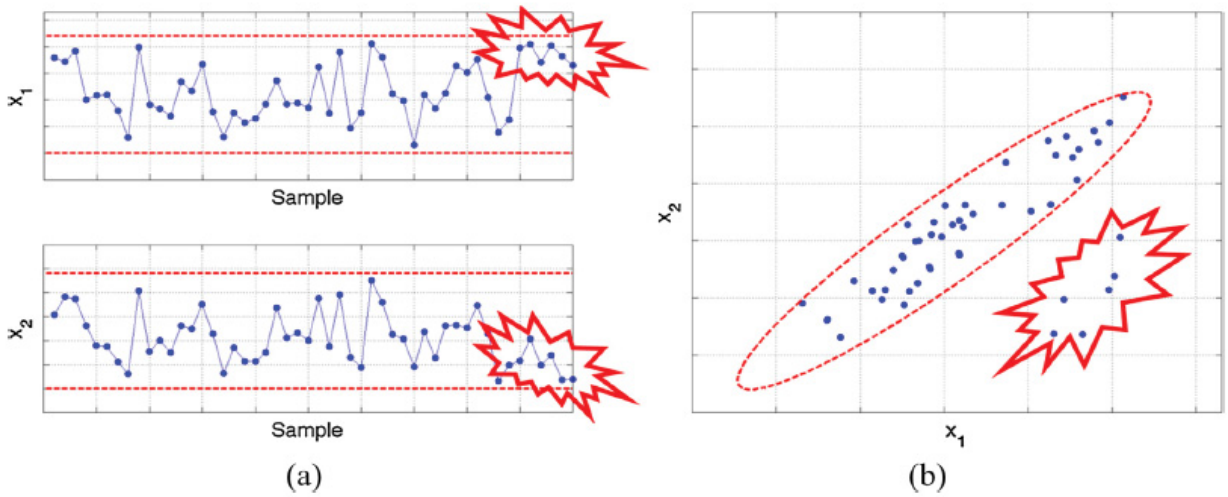
Conforme Philip Chen e Zhang (2014), a estatística está relacionada à coleta, organização e interpretação dos dados. Tais técnicas são utilizadas para analisar correlações e relações entre diferentes tipos de dados, além de fornecer valores numéricos acerca dos resultados. No âmbito da análise de grandes volumes de dados, as análises estatísticas tradicionais não são as mais adequadas para serem utilizadas, sendo neste caso buscado por outras abordagens complementares. Como exemplo da aplicação estatística em análise de dados, Olshannikova et. al. (2015) comentam que as técnicas estatísticas podem ser aplicadas em modelos preditivos. Nestes modelos a regressão é utilizada para determinar alterações entre variáveis dependentes e independentes. Em outras situações a regressão linear pode ser aplicada para verificar a interdependências entre as variáveis, buscando resultados explicativos e explorá-los para fazer previsões.

2.3.1.7.1 Utilização de análise multivariada em modelos de análise

Segundo He e Wang (2017), para tratar os desafios de *Big Data* são necessárias abordagens multivariadas, pois no processo de análise de dados com o uso da estatística univariada foram observadas diversas dificuldades. Conforme Junqué de Fortuny et. al. (2013), para a aplicação de análise preditiva, o modelo multivariado é superior ao modelo de eventos univariado.

Como exemplo, nos gráficos (a) e (b) da FIGURA 11, as propriedades de $\boldsymbol{X_1}$ e $\boldsymbol{X_2}$ foram medidas para um produto. Neste exemplo, os gráficos univariados representados por (a) não conseguem detectar as falhas ocorridas nas últimas seis amostras, porém o exemplo (b) representando um gráfico multivariado, mostra claramente que as medições de $\boldsymbol{X_1}$ e $\boldsymbol{X_2}$ são correlacionadas positivamente, enquanto que as últimas seis amostras não seguem tal correlação positiva. (HE e WANG, 2017).

FIGURA 11 – COMPARAÇÃO ENTRE MODELOS UNIVARIADOS E MULTIVARIADOS



FONTE: HE e WANG (2017).

2.4 ANÁLISE MULTIVARIADA

Perante um conjunto de pequenos objetos, a estatística clássica firmou-se no estudo de uma única característica, ou variável, desenvolvendo assim noções de testes e estimativas firmadas em hipóteses muito restritivas. Contudo, em observações práticas, os objetos são caracterizados por um grande número de características, ou variáveis. (BOUROUCHE e SAPORTA, 1982).

Nas áreas de pesquisa, a análise multivariada é amplamente empregada, sendo utilizada em casos em que os dados de um ou mais objetos relevantes são obtidos por múltiplas variáveis de interesse. Desta forma, toda amostra que dispõe de duas ou mais variáveis de interesse pode ser considerada como multivariada. A medição, interpretação e previsão do grau de relacionamento das variáveis estatísticas é uma das principais finalidades da análise multivariada. Desta forma, a natureza multivariada reside nas combinações múltiplas de variáveis estatísticas e não unicamente no número de observações e variáveis. (HAIR, ANDERSON, *et al.*, 2005).

Segundo Johnson e Winchern (2007), os objetivos de pesquisa mais aplicados à análise multivariada são:

I. Redução de dados ou simplificação estrutural, tornando os dados mais simples possíveis para serem estudados e interpretados, porém sem descartar nenhuma informação significativa para o estudo;

II. Ordenação e agrupamento, criação de regras para classificação de grupos de objetos ou variáveis pelas suas características ou semelhanças;

III. Investigação da dependência entre as variáveis, observando as suas independências ou dependências em relação às variáveis de interesse;

IV. Predição, observando a interação entre os dados, com o objetivo de prever situações baseadas na relação entre as variáveis;

V. Construção de hipóteses e testes, com o objetivo de validar suposições ou convicções a respeito dos dados, baseadas em formulação de hipóteses estatísticas.

2.4.1 Conceitos de Análise Multivariada

Com o objetivo de estudar um fenômeno, a análise multivariada é um método estatístico modelado para obter informações a partir de um conjunto de dados gerados ou medidos de muitas variáveis. (JOHNSON e WICHERN, 2007).

Segundo Lattin et. al. (2011), a aplicação de métodos multivariados para análise de dados tem aumentado significativamente, isso se deve principalmente a dois fatores:

I. Cada vez mais os pesquisadores se deparam com a complexidade do comportamento humano, o qual pode ser exposto por meio de dados multivariados;

II. O avanço ocorrido nas tecnologias da informação, como o *Big Data*, e que tem gerado imensa quantidade de dados, não apenas do comportamento humanos, mas também das interações ao seu redor.

O grande paradoxo não é a imensa quantidade de dados, ou em alguns casos a sua ausência, mas sim a capacidade e habilidade de extração, análise e compreensão destes dados. Neste momento que os métodos multivariados cumprem a sua função. (LATTIN, CARROLL e GREEN, 2011).

Para Hair et al. (2005), “o propósito da análise multivariada é medir, explicar e prever o grau de relacionamento entre as variáveis, de modo que o caráter multivariado é constituído por combinações múltiplas de variáveis”.

2.4.2 Algumas definições sobre análise multivariada

As análises multivariadas surgem da necessidade de um pesquisador entender um fenômeno por meio de dados que são compreendidos por um número de variáveis p>1. Os valores destas observações são atribuídos para cada item distinto, e sua representação é dada pela notação X_{jk} para indicar o j-ésimo valor de unidade amostral e da k-ésima variável medida. (JOHNSON e WICHERN, 2007).

Segundo Lattin et al. (2011), a análise multivariada pode ser definida como:

> Os métodos multivariados são definidos como um conjunto de procedimentos para analisar associação entre dois ou mais conjuntos de medidas que foram feitas em cada objeto em uma ou mais amostras de objeto. Caso apenas dois conjuntos de medidas estejam envolvidos, os dados são referidos como bivariados. (LATTIN, et al., 2011, p.3).

Tais conjuntos de medidas são organizados em estruturas denominadas de matrizes, conforme TABELA 2, as quais representam em suas em suas linhas o número de objetos da amostra (n) e em suas colunas o número de variáveis (p).

TABELA 2 – MATRIZ DE AMOSTRAS POR VARIÁVEIS

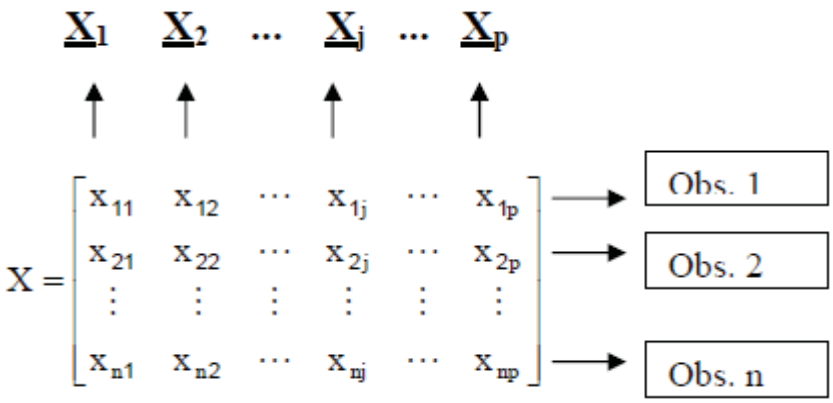
Unidades amostrais ou experimentais	Variáveis			
	1	2 ...	K ...	p
1	X_{11}	X_{12} ...	X_{1K} ...	X_{1P} ...
2	X_{21}	X_{22} ...	X_{2K} ...	X_{2P} ...
⋮	⋮	⋮	⋮	⋮
j	X_{J1}	X_{J2} ...	X_{JK} ...	X_{JP} ...
⋮	⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2} ...	X_{nK} ...	X_{nP} ...

FONTE: Adaptado JOHNSON e WICHERN (2007).

Segundo Marques (2005), uma observação multivariada é um conjunto de valores de p variáveis distintas a respeito de um mesmo item, onde tais dados podem

ser distribuídos em uma matriz $\boldsymbol{X}$ de ordem $n\boldsymbol{X}_p$, conforme FIGURA 12. Desta forma, a matriz $\boldsymbol{X}$ equivale a uma amostra de tamanho n proveniente de uma população p-variada, onde cada linha de $\boldsymbol{X}$ é uma observação multivariada.

FIGURA 12 – OBSERVAÇÃO MULTIVARIADA EM UMA MATRIZ X



FONTE: MARQUES (2017).

Segundo Hair et. al. (2005), para compreender os conceitos da abordagem multivariada, são expostos alguns conceitos importantes, como variável estatística, escalas de medida, dados qualitativos e quantitativos, e erros de medidas, os quais podem ser melhor descritos na sequência.

2.4.2.1 Variável estatística

É uma combinação linear de variáveis determinadas pelo analista de dados, e com pesos determinados por técnicas multivariadas. Uma variável estatística de X_1 a X_n pode ser descrita por meio da equação (1.8).

$$Valor\ da\ variável\ estatística = w_1X_1 + w_2X_2 + w_3X_3 + \cdots + w_RX_n \qquad (1.8)$$

Onde X_n é a n-ésima variável observada e W_n é o n-ésimo peso determinado pela técnica multivariada. Como resultado, se obtém um único valor, que dependendo da técnica empregada, captura o melhor caráter multivariado da análise. (HAIR, ANDERSON, *et al.*, 2005).

2.4.2.2 Escalas de medidas

O conceito de medida é importante para que o analista de dados possa escolher a técnica multivariada apropriada e representar com precisão o conceito de interesse. Neste caso há dois tipos básicos de dados, os não métricos ou qualitativos, e os métricos, ou quantitativos. (HAIR, ANDERSON, *et al.*, 2005).

2.4.2.3 Dados qualitativos

Tratam de características, propriedades ou atributos que descrevem ou identificam um objeto. Podem ser diferenciadas de duas formas, nominal ou ordinária.

Nas escalas nominais, também conhecidas como escalas categóricas, são designados números para identificar indivíduos ou objetos, porém tais números não possuem significado quantitativo, apenas indicação de presença ou ausência da característica. (HAIR, ANDERSON, *et al.*, 2005).

Nas escalas ordinais, as variáveis são ordenadas ou ranqueadas de acordo com a quantidade de atributos que possui. Também neste caso, os números utilizados nas escalas ordinais, não são quantitativos, pois são apenas indicações de uma ordenação. (HAIR, ANDERSON, *et al.*, 2005).

2.4.2.4 Dados quantitativos

Especificam por meio de um valor numérico as características de identificação ou diferenciação de indivíduos, circunstâncias ou objetos. São compostos por duas escalas que fornecem altos níveis de precisão de medida, as escalas intervalares e as escalas de razão. As escalas intervalares. Possuem um ponto zero arbitrário, e as escalas de razão possuem um ponto zero absoluto. Um exemplo de escala intervalar são as escalas termométricas Fahrenheit e Celsius, pois cada uma possui um ponto zero arbitrário diferenciado, além de que nenhuma delas possui quantidade nula de temperatura ou valores nulos. Nas escalas de razão todas as operações matemáticas são possíveis, pois todas elas apresentam um zero absoluto, o que permite elevada precisão de medida.

O conhecimento da existência e aplicação dos dois diferentes tipos de escala é altamente justificado para que o analista de dados não utilize de forma inadequada dados quantitativos em lugar de qualitativos e vice-versa. Outra razão, e de suma

importância para a análise de dados, é em relação à escolha da técnica multivariada mais adequada para ser aplicada aos tipos de dados. (HAIR, ANDERSON, *et al.*, 2005).

2.4.2.5 Erros de medida

São considerados como erros de medidas os valores que apresentados não representam os valores verdadeiros. Estes valores possuem variação, alguns durante a entrada de dados, devido a problemas de imprecisão, como por exemplo, medir uma escala de valores com cinco variáveis, quando tem-se a possibilidade de apenas responder a três valores. Outras imprecisões em valores podem ser ocasionadas também pela falta de habilidade em retornar informações precisas, como por exemplo, questionamentos sobre uma fonte de renda podem ser corretos, mas pouco precisos. Devido a isso, as técnicas utilizadas pelas abordagens multivariadas devem considerar certo grau de erro de medida, sendo evidenciados ruídos nas variáveis observadas, e devendo ser considerado nos resultados tanto valores verdadeiros, quanto ruídos. (HAIR, ANDERSON, *et al.*, 2005).

Segundo Hair et. al. (2005), durante o processo o analista de dados pode avaliar o nível de erro presente e levar em consideração a validade e a confiabilidade dos dados. A validade é a precisão de representatividade de uma medida e a confiabilidade é o quanto a variável observada está livre de erros e retornando valores verdadeiros, sendo antagônico ao erro de medida.

Reduzir os erros de medida podem melhorar os resultados obtidos, porém apesar de ser um processo complexo e que demanda certo grau de esforço, o analista de dados deve sempre trabalhar tendo em mente a busca por dados validos e os mais confiáveis possíveis, o que resultarão em um resultado mais fidedigno em relação às variáveis de interesse. (HAIR, ANDERSON, *et al.*, 2005).

2.4.3 Classificação das técnicas multivariadas

Segundo Lattin et al. (2011), para a aplicação de métodos multivariados é levado em consideração à natureza dos diferentes tipos de dados. Para tanto é necessário observar três características das análises:

I. Se a técnica a ser utilizada é aplicada para a análise da dependência ou interdependência dos dados;

II. Se a técnica a ser utilizada é aplicada para a exploração ou com o objetivo de confirmação dos dados;

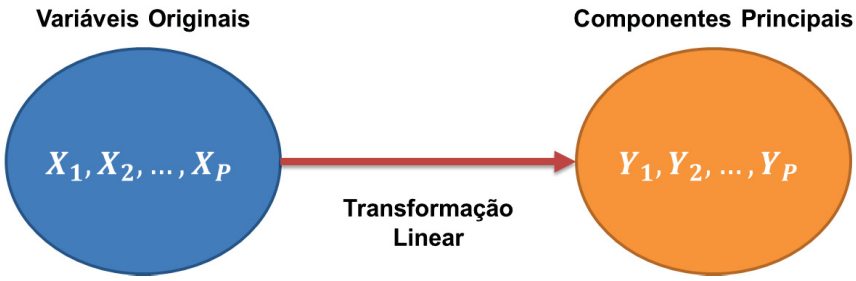
III. Se a técnica a ser utilizada é aplicada com dados métricos (que podem ser suscetíveis a operações aritméticas) ou com dados não métricos (que podem ser descritos ou ordenados).

2.5 ANÁLISE DE COMPONENTES PRINCIPAIS

Os pesquisadores se deparam com inúmeras variedades e quantidades de dados durante as suas pesquisas. Tais dados possuem diferentes dimensões e características, e desta forma, com tantas variáveis, se torna complicado de entender, incorporar e até mesmo visualizar os padrões e associações entre elas. Além disso, este processo se torna complexo, pois existe considerável redundância entre os dados, o que gera altos níveis de correlação e multicolinearidade. (LATTIN, CARROLL e GREEN, 2011).

Neste cenário, a análise de componentes principais torna-se uma importante técnica para auxiliar na compreensão e possui como objetivo analisar um conjunto de dados mutivariados, com p variáveis correlacionadas, é recomendado transformar este conjunto de dados originais em um novo conjunto de variáveis não correlacionadas, denominado de componentes principais. Estes componentes principais são ordenados de forma decrescente de importância, por meio das combinações lineares das variáveis originais, de maneira que o primeiro componente principal seja o que possuam a máxima variância, e assim sucessivamente, como resumo apresentado na FIGURA 13. (MARQUES, 2005).

FIGURA 13 – RESUMO DO PROCESSO DE COMPONENTES PRINCIPAIS



FONTE: Adaptado MARQUES (2017).

Segundo Johnson e Winchern (2007), por meio de poucas combinações lineares, a técnica de análise de componentes principais procura a explicação de uma estrutura de variância-covariância de um conjunto de variáveis. Os objetivos originais de tal análise estão relacionados com:

I. Redução da dimensão de dados originais;

II. Melhoria na interpretação dos dados analisados.

Mesmo precisando de p elementos para explicar a variabilidade de um sistema, em muitos casos uma grande parcela desta variabilidade pode ser explicada por um pequeno número k de componentes principais. Desta forma entende-se que sendo $k \leq p$, existe tanta informação nos componentes k, quanto nas variáveis p, podendo assim as variáveis originais p serem substituídas pelos componentes principais k. (JOHNSON e WICHERN, 2007).

> Sempre que o tamanho do conjunto de dados torna-se difícil de manejar (em termos do número de variáveis), os componentes principais podem ser úteis na redução dessa dimensionalidade. Trabalhar com menos dimensões torna mais fácil visualizar os dados e identificar padrões interessantes. (LATTIN, et al., 2011, p.68).

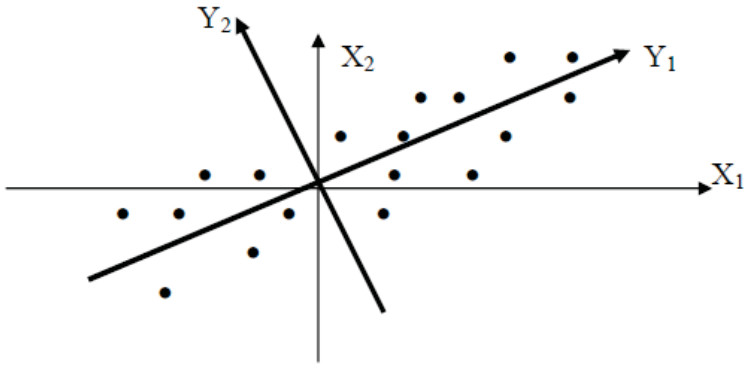
Ainda, segundo Johnson e Winchern (2007), a análise de componentes principais não é considerada como uma técnica que é empregada para se obter um resultado conclusivo, pois em muitos casos ela destina-se a uma investigação intermediária para a aplicação posterior de outras abordagens multivariadas mais complexas.

Para Lattin et al. (2011), cada componente principal é definido pelas variáveis com as quais é mais altamente correlacionada e Johnson e Winchern (2007) complementa que algebricamente, as componentes principais são uma combinação linear das variáveis aleatórias p $(X_1, X_2, \dots, X_p)$.

Geometricamente tais combinações caracterizam um novo sistema de coordenadas, que são obtidos pela rotação do conjunto de dados originais $X_1, X_2, \dots, X_p$ como eixos. Como consequência da rotação, $Y_1, Y_2, \dots, Y_p$ são os novos eixos obtidos e que representam as direções de máxima variabilidade, conforme FIGURA 14,

trazem uma explicação mais simples da estrutura de covariância. (JOHNSON e WICHERN, 2007).

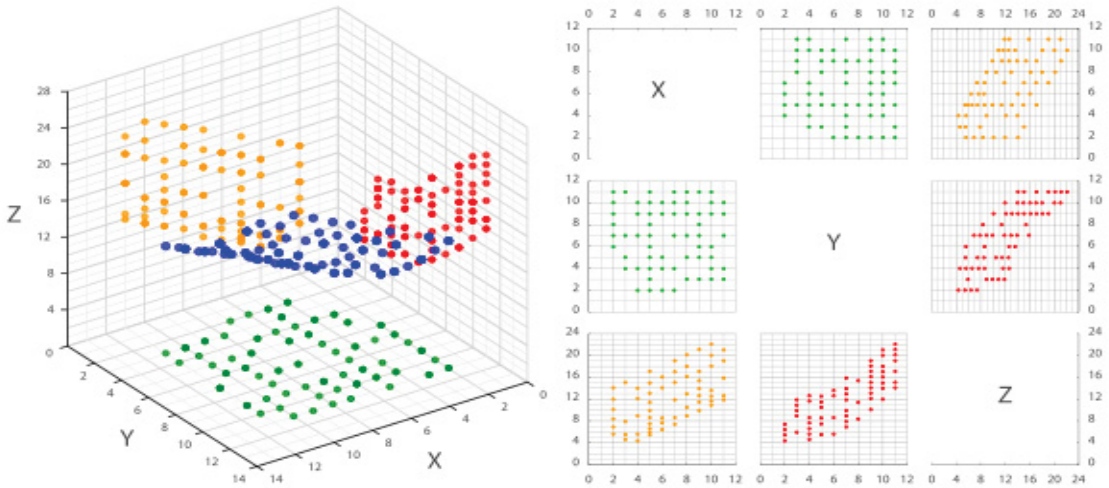
FIGURA 14 – EXEMPLO DA REPRESENTAÇÃO GEOMÉTRICA PARA $P = 2$



FONTE: MARQUES (2017).

Os componentes principais possuem a propriedade de que cada componente não está correlacionado com os demais, o que beneficia de forma positiva o processo devido à eliminação da multicolinearidade quando se utiliza a técnica em uma análise de dependência, como por exemplo, em uma análise de regressão. Geometricamente, conforme demonstrado na FIGURA 15, a variância máxima e a combinação linear são respectivamente a medida de dispersão dos dados e a projeção destes dados em pontos em um único eixo no espaço tridimensional. (LATTIN, CARROLL e GREEN, 2011).

FIGURA 15 – DIAGRAMA DE DISPERSÃO DOS VALORES X_1, X_2 *e* X_3



FONTE: WIKIPEDIA (2018).

2.5.1 Processo para análise de componentes principais

Segundo Manly (2008), a análise de componentes principais inicia-se com a declaração dos dados de p variáveis para p indivíduos, como demonstrado na TABELA 3.

TABELA 3 – FORMA DOS DADOS PARA UMA ANÁLISE DE COMPONENTES PRINCIPAIS

Caso	Variáveis			
	X_1	X_2	...	X_P
1	X_{11}	X_{12}	...	X_{1P}
2	X_{21}	X_{22}	...	X_{2P}
⋮	⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	...	X_{nP}

FONTE: Adaptado MANLY (2008).

Assim, de acordo com Johnson e Winchern (2007), os p componentes principais $(Y_1, Y_2, \dots, Y_p)$ são combinações lineares dadas por (1.9) quando existe um vetor aleatório $\boldsymbol{X} = (Y_1, Y_2, \dots, Y_p)$ originário de uma população com variância Σ, e com autovalores $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

$$
\begin{aligned}
Y_1 &= \boldsymbol{a}'_1\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
Y_2 &= \boldsymbol{a}'_2\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \qquad \vdots \qquad\qquad \vdots \\
Y_p &= \boldsymbol{a}'_p\boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned}
\tag{1.9}
$$

Pode ser deduzido que Y_1, que representa a primeira componente principal, possui a maior variância de todos as demais componentes, e que quanto maior esse valor, maior o número de informações disponíveis a respeito dos dados originais

estará contido neste único componente, porém mesmo que Y_1 explique uma grande parcela da variação dos dados originais, ele não explica toda a variação. A fração residual não explicada por Y_1 é tratada pela segunda componente principal Y_2, a qual explica o valor máximo das variações ainda não explicadas por Y_1. As próximas frações residuais serão explicadas pelo componente Y_3 e assim sucessivamente. Os componentes principais $Y_1, Y_2, \dots, Y_p$ possuem a propriedade de serem mutuamente não correlacionados, ou seja, os componentes principais subsequentes são determinados para não serem correlacionados aos componentes principais anteriores, além de cada novo componente principal ser destinado a explicar a máxima quantidade possível de variação ainda não explicada. (LATTIN, CARROLL e GREEN, 2011).

Assim, Johnson e Winchern (2007), definem sucintamente que:

- O primeiro componente principal é a combinação linear $\boldsymbol{a'}_1\boldsymbol{X}$ que maximiza $Var\ (\boldsymbol{a'}_1\boldsymbol{X})$ sujeito a $\boldsymbol{a'}_1\boldsymbol{a}_1 = 1$
- O segundo componente principal é a combinação linear $\boldsymbol{a'}_2\boldsymbol{X}$ que maximiza $Var\ (\boldsymbol{a'}_2\boldsymbol{X})$ sujeito a $\boldsymbol{a'}_2\boldsymbol{a}_2 = 1$ e $Cov\ (\boldsymbol{a'}_1\boldsymbol{X}, \boldsymbol{a'}_2\boldsymbol{X}) = 0$
- Nas próximas etapas, o componente principal Y_i é a combinação linear $\boldsymbol{a'}_i\boldsymbol{X}$ que maximiza $Var\ (\boldsymbol{a'}_i\boldsymbol{X})$ sujeito a $\boldsymbol{a'}_i\boldsymbol{a}_j = 1$ e $Cov\ (\boldsymbol{a'}_i\boldsymbol{X}, \boldsymbol{a'}_k\boldsymbol{X}) = 0$ para $k < i$.

Segundo Manly (2008), a ausência de correlações demonstra que os índices estão retornando diferentes dimensões de dados, conforme equação (1.10),

$$Var\ (Y_1)\ \geq Var\ (Y_2)\ \geq \cdots \geq Var\ (Y_p) \tag{1.10}$$

A matriz de covariância é simétrica e representada por (1.11), onde o elemento S_{ii} na diagonal é a variância de X_i, e os termos não pertencentes a diagonal S_{ij} são a covariância entre as variáveis X_i e X_j.

$$\boldsymbol{\Sigma} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \tag{1.11}$$

Os autovalores da matriz C representam as variâncias dos componentes principais. Existem p autovalores, alguns podendo ser zero, porém de forma alguma apresentar valores negativos, onde λ_i corresponde ao i-ésimo componente principal conforme (1.12).

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \tag{1.12}$$

Devido à análise de componentes principais buscar a máxima variância, ela pode ser vulnerável as diferenças desnecessárias de escalas entre as variáveis. Em virtude desta situação é recomendado padronizar os dados para que possuam média igual a zero e variância igual a um. A solução para tal questão é a aplicação de uma decomposição de autovalor da matriz de correlação, que é a matriz de covariância dos dados padronizados. (LATTIN, CARROLL e GREEN, 2011).

- Desta forma cada autovetor representa a direção de um dos eixos principais;
- Cada autovalor (λ) é igual a variância do componente principal Y_i, definido por $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$;
- A matriz de covariância estimada dos componentes principais, representada por $\boldsymbol{S}$, é uma matriz diagonal com $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$;

A análise de componentes principais para dados padronizados é realizada por meio da matriz de correlação $\boldsymbol{R}$, representada em (1.13).

$$\boldsymbol{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \tag{1.13}$$

Analisando os componentes principais estimados, buscam-se variáveis com baixos coeficientes nos componentes, tendo como objetivo o descarte dessas. Outra maneira, para a análise dos dados, é utilizar apenas os primeiros e mais significativos componentes principais, desde que a soma de suas variâncias seja um percentual representativo perante o total de todos os p componentes. (MANLY, 2008).

Assim se os k componentes principais, sendo $k < p$, explicarem em torno de 80 a 90% da variabilidade dos dados, então se pode atribuir aos primeiros componentes a representação das p variáveis originais sem uma perda significativa de informações. (JOHNSON e WICHERN, 2007).

Existem diversas técnicas para determinar o número de componentes principais para análise, em muitas delas se faz necessário obter uma quantidade suficiente de componentes principais para explicar adequadamente a variância em cada variável original. (LATTIN, CARROLL e GREEN, 2011).

Segundo Moriggi (2018), uma das técnicas utilizadas para determinar a quantidade de componentes principais necessária para uma análise é denominada método de Jolliffe, a qual aplica critérios específicos para a seleção das variáveis. De acordo com Jolliffe (1972), a seleção de variáveis a partir da análise de componentes principais é baseada nos valores que foram obtidos pelos autovalores, sendo para isso utilizada duas metodologias chamadas de B2 e B4, onde as mesmas buscam excluir m variáveis a partir de um critério de seleção que segrega as variáveis com valores menores ou iguais a 0,7.

A metodologia B2 consiste em associar cada uma das variáveis a cada um dos componentes principais por meio de uma matriz de correlação. Para cada componente principal é verificada a variável que possui maior valor de correlação, sendo selecionada a variável de maior correlação com o último componente principal, posteriormente a próxima variável de maior correlação é associada a penúltima componente principal, e assim sucessivamente. Determinando tais variáveis, estas são descartadas, sendo selecionado neste caso as variáveis que não possuíam alto valor de correlação com os últimos componentes principais. (JOLLIFFE, 1972).

A metodologia B4 é semelhante e emprega o mesmo critério da B2, realizando a correlação das variáveis com cada um dos componentes principais, sendo selecionadas as variáveis que possuem maior correlação com cada componente principal, porém neste caso, de forma oposta a B2, o método é iniciando pela primeira componente, seguida da segunda e assim sucessivamente, sendo posteriormente descartadas as demais as variáveis que não foram selecionadas. (JOLLIFFE, 1972).

Tanto em Jolliffe B2, quanto em Jolliffe B4, caso uma mesma variável apresente a maior correlação em mais de um componente principal, é selecionada a próxima variável com maior correlação da componente correspondente, e assim

sucessivamente enquanto existirem autovalores iguais ou menores que 0,7, pois componentes principais com baixos autovalores não geram impacto significativo na variabilidade dos dados analisados, e a eliminação das variáveis de alta correlação com esses componentes não gera perda de informações para o modelo. (JOLLIFFE, 1972).

Segundo Moriggi (2018), é possível observar que os métodos de Jolliffe B2 e B4 são complementares, pois as variáveis que são mantidas pelo método de Jolliffe são iguais ao número de autovalores maiores que 0,7. Tal método se destaca por selecionar, por meio das primeiras componentes principais, as variáveis mais significantes para o modelo.

2.6 REGRESSÃO LINEAR MÚLTIPLA

Segundo Lattin et. al. (2011), a análise de regressão múltipla possui como um dos objetivos o entendimento da relação entre um conjunto de variáveis independentes (X) e uma única variável dependente (Y), por meio de uma combinação linear, onde as variáveis independentes se correspondam de forma próxima a variável dependente. A regressão múltipla é uma das técnicas mais utilizadas para análise de dependência, sendo usada:

- Na descrição de relação de forças do modelo, entre as variáveis independentes e a variável dependente.
- Na inferência estatística, verificando a relação de significância descrita pelo modelo.
- Na geração de modelos preditivos, por meio das variáveis independentes para obter uma previsão da variável dependente.

Segundo Levine, Berenson e Stephan (1998), com o propósito de desenvolver um modelo estatístico preditor, a análise de regressão possui muita aplicação com o objetivo de previsão, utilizando as variáveis independentes para obter uma variável resposta. A equação utilizada é derivada de uma forma mais básica de regressão, a linear simples, o qual segundo Marques e Marques (2009) pode ser demonstrada pela Equação (1.14):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{1.14}$$

Na equação (1.14), β_0 é denominado de intercepto, β_1 de coeficiente de regressão linear, e ε_i de erro aleatório para a observação i.

Conforme Levine, Berenson e Stephan (1998), quando existe mais de uma variável explicativa para prever o valor da variável dependente, e podendo existir p variáveis independentes, o modelo de regressão linear simples pode ser estendido, considerando apenas combinações lineares das variáveis independentes. Para Montgomery e Runger (2009), em muitas análises de regressão existem situações com a presença de mais de um regressor, sendo este modelo denominado de modelo de regressão múltipla. Tal modelo pode ser escrito conforme equação (1.15).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_p X_{pi} + \varepsilon_i \quad (1.15)$$

Na equação de regressão múltipla (1.15) os elementos representam:

- β_0 = interseção de Y;
- β_1= inclinação de Y em relação a variável X_1, mantendo constante X_2, X_3, ... , X_p;
- β_2= inclinação de Y em relação a variável X_2, mantendo constante X_1, X_3, ... , X_p;
- β_3= inclinação de Y em relação a variável X_3, mantendo constante X_1, X_2, ... , X_p;
- β_p= inclinação de Y em relação a variável X_p, mantendo constante X_1, X_2, ... , X_{p-1};
- ε_i = erro aleatório em Y para a observação i;

O parâmetro β_0 é onde ocorre a interseção com o plano. O elemento β_1 é um dos coeficientes de regressão, pois β_1 mede as variações ocorridas em Y devido às variações ocorridas em X_1, enquanto as demais variáveis são mantidas constantes, e desta forma sucessivamente. Assim, o modelo é constituído de p variáveis regressoras, ou independentes (X), as quais podem estar relacionadas a variável dependente (Y). (MONTGOMERY e RUNGER, 2009).

Segundo Montgomery e Runger (2009), em casos de grandes quantidades de variáveis, pode ser útil utilizar expressões matriciais para expressar o modelo de regressão linear múltipla. Sendo um modelo de n equações, o mesmo pode ser expressado de forma matricial como a equação (1.16):

$$Y = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1.16}$$

Onde:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.6.1 Coeficiente de determinação múltipla - R^2

Após a conclusão do modelo de regressão linear múltiplo, é possível calcular o denominado coeficiente de determinação (R^2), o qual representa o percentual de explicação da variável dependente (y) em relação as variáveis explicativas (X). (LEVINE, BERENSON e STEPHAN, 1998).

Segundo Lattin et. al. (2011), R^2 pode ser usado para verificar a concordância do modelo, demonstrando a proporção de incerteza da variável dependente, e representando a qualidade de ajuste desta em relação ao modelo de regressão, sendo R^2 representado pela Equação (1.17).

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - y)^2} \tag{1.17}$$

Quando uma variável é adicionada ao modelo, a estatística R^2 acaba por aumentar simultaneamente. Devido a tal situação é recomendada a utilização do R^2 ajustado, o qual aumentará apenas se a variável adicionada ao modelo reduzir a medida quadrática do erro. O R^2 ajustado auxilia a preservar o modelo em relação a ajustes desnecessários ou em excesso, devido a inclusão de variáveis independentes que não são úteis. (MONTGOMERY e RUNGER, 2009).

Para que exista conjunturas em relação ao modelo, devido ao número de variáveis intendentes utilizados pela amostra, é sugerido a utilização do R^2 ajustado, sendo recomendado principalmente quando existe comparação entre dois ou mais modelos que trabalham com a mesma variável de resposta, mas que possuem diferentes quantidades de variáveis explicativas. (LEVINE, BERENSON e STEPHAN, 1998).

A equação (1.18) representa o R^2 ajustado, sendo P o número de variáveis independentes do modelo de regressão.

$$R^2_{ajust.} = 1 - \left[(R^2)\frac{n-1}{n-P-1}\right]$$

$$R^2_{ajust.} = 1 - \left[\left(1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - y)^2}\right)\frac{n-1}{n-P-1}\right] \quad (1.18)$$

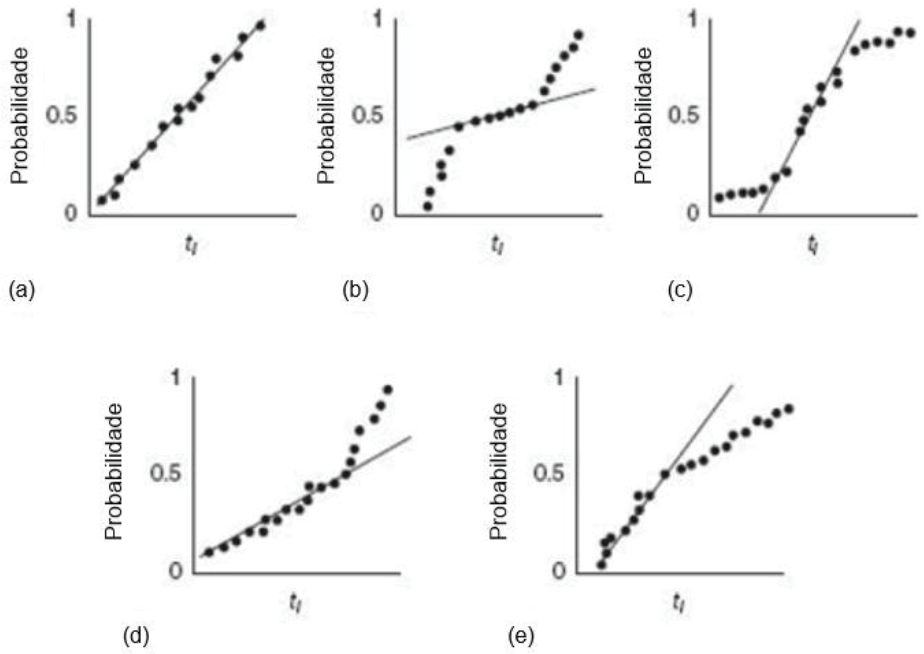
2.6.2 Análise de resíduos na regressão linear múltipla

A análise de resíduos é utilizada para verificar se o modelo de regressão se ajustou adequadamente aos dados. Tais resíduos, ou também denominados de erros estimados (e_i), representam a diferença entre os valores obtidos da variável dependente (Y_i), pelos valores previstos ($\hat{Y}_i$) em relação aos dados amostrais, conforme demonstrado na equação (1.19).

$$e_i = Y_i - \hat{Y}_i \quad (1.19)$$

Segundo Montgomery e Runger (2009), a análise dos resíduos gerados pela regressão múltipla permite a avaliação dos resultados da regressão, sendo frequentemente úteis para observar o comportamento do modelo, muitas vezes realizado de forma gráfica para analisar o padrão de comportamento do modelo e em que este pode ser melhorado. A análise de resíduos é composta pelo teste de normalidade, homocedasticidade e independência dos resíduos. Todas as análises podem ser feitas de forma visual, por meio de gráficos, ou por testes de hipóteses específicos para cada situação.

FIGURA 16 – GRÁFICO PARA VERIFICAÇÃO DA NORMALIDADE DOS RESÍDUOS



FONTE: MONTGOMERY, PECK e VINING (2012).

Na FIGURA 16 é possível observar os gráficos de verificação de normalidade para cinco situações distintas, conforme abaixo:

a) Distribuição normal ideal;
b) Distribuição com cauda leve (*light-tailed distribution*);
c) Distribuição com cauda pesada (heavy-tailed distribution);
d) Distribuição com inclinação positiva;
e) Distribuição com inclinação negativa;

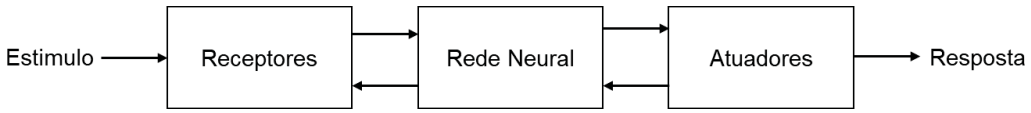
Para as situações (b), (c), (d) e (e) a normalidade não é verificada. Para se testar a normalidade dos resíduos pode-se utilizar testes de hipóteses tais como, Anderson-Darling, Shapiro-Wilk e Kolmogorov-Smirnov. Para se testar a homocedasticidade (variâncias constantes) dos resíduos utiliza-se o teste de hipótese de Breusch-Pagan e para o teste da independência dos resíduos utiliza-se o teste de Durbin-Watson. (MARQUES, 2005).

2.7 REDES NEURAIS ARTIFICIAIS (RNA)

Segundo Haykin (2001), as Redes Neurais Artificiais (RNA) têm sido estudadas desde a percepção da capacidade de processamento diferenciada do cérebro

humano. Esta capacidade é um diferencial em relação aos computadores convencionais, pois o cérebro é um sistema altamente complexo, constituído de estruturadas denominadas de neurônios, os quais possuem a capacidade de processar uma grande quantidade de informações de forma não-linear e em paralelo. O sistema nervoso humano possui três estágios, conforme FIGURA 17. O primeiro estágio é o que recebe os estímulos do ambiente e os converte em informações para o cérebro (rede neural). Os atuadores convertem os sinais enviados pela rede em respostas, como saídas do sistema. Dentro desse sistema existem os neurônios, os quais são a unidade de processamento de informações e que são fundamentais para o funcionamento de uma rede neural. As sinapses são as estruturas elementares que realizam a interação entre os neurônios.

FIGURA 17 – REPRESENTAÇÃO EM DIAGRAMA DE BLOCOS DO SISTEMA NERVOSO



FONTE: Adaptado HAYKIN (2001).

Os sistemas de RNAs se baseiam no comportamento e funcionamento dos neurônios biológicos, principalmente em relação as unidades de computação paralelas e distribuídas de forma que se comunicam por meio de conexões sinápticas. (BRAGA, CARVALHO e LUDERMIR, 2011).

Segundo Haykin (2001), as RNAs foram projetadas para recriar a forma como o cérebro executa uma determinada função, sendo possível descrever a seguinte definição para uma RNA.

> Uma rede neural é um processador maciçamente paralelamente distribuído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos: (HAYKIN,2001, p.28).
>
> 1. O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
>
> 2. Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

A utilização de Redes Neurais Artificiais traz algumas propriedades úteis, dentre elas:

- Não-linearidade: Uma rede neural artificial pode ou não ser linear, porém essa é uma importante propriedade, principalmente se os dados de entrada não forem lineares;
- Mapeamento de entrada-saída: O aprendizado de uma RNA está relacionado ao mapeamento de entrada-saída dos dados, os quais são repetidos diversas vezes até que se alcance um estado estável em que não existam mais modificações significativas;
- Adaptabilidade: Uma RNA já treinada em um determinado ambiente, pode ser retreinada para se adaptar a pequenas modificações existentes neste meio, sendo esta uma propriedade útil das redes neurais;
- Resposta a evidências: Uma rede neural pode ser desenvolvida não apenas para fornecer informações sobre padrões, mas também para demonstrar a confiança a respeito de uma decisão tomada;
- Informação contextual: Cada neurônio da rede é afetado pela atividade dos demais neurônios contidos nesta rede, e consequentemente, as informações contextuais são tratadas naturalmente pela RNA;
- Tolerância a falhas: Uma rede projetada para utilização computacional, possui o potencial de tolerância a falhas do sistema, devido a um projeto de computação robusta, sendo que sua performance pouco se degrada em condições adversas de funcionamento.

Sendo compostas por unidade de processamento simples, as RNAs calculam funções matemáticas normalmente não-lineares. Tais sistemas são dispostos em paralelo, distribuídas em uma ou mais camadas interligadas por uma grande quantidade de conexões, sendo estas associadas a pesos, os quais armazenam as informações apresentadas no modelo, tendo como função a ponderação da entrada recebida nos neurônios da rede. (BRAGA, CARVALHO e LUDERMIR, 2011).

Segundo Braga et. al. (2011), a utilização da RNA é atrativa, pois fornecem um desempenho superior a criação de modelos, se comparado aos métodos matemáticos convencionais. A rede neural artificial, inicia-se com uma fase de aprendizagem, onde um conjunto de exemplos são demonstrados para a rede, sendo extraído posteriormente informações que representam os dados fornecidos inicialmente, sendo estes resultados utilizados para gerar soluções aos problemas apresentados.

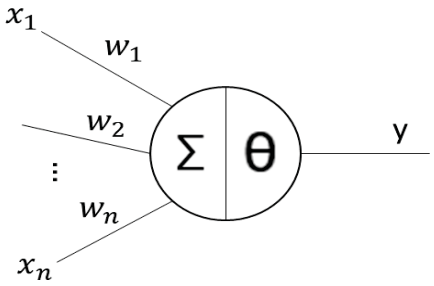
A utilização das redes neurais artificiais vai além de mapear entradas e saídas de dados. O diferencial das RNAs, e o seu atrativo diferencial é a sua capacidade de aprendizado por meio dos exemplos apresentados, sendo capaz de generalizar estas informações e aprender por meio de um conjunto reduzido de informações, gerando

posteriormente respostas para um conjunto de dados não conhecidos, sendo capazes de fornecer informações não explícitas por meio de exemplos. Em termos de funções multivariadas, as RNAs conseguem mapear tais funções, tendo considerável capacidade de auto-organização e de processamento temporal, tornando a RNA como uma alternativa computacional para a solução de problemas complexos e não-lineares. (BRAGA, CARVALHO e LUDERMIR, 2011).

2.7.1 Os neurônios artificiais

Com base nos neurônios biológicos, no ano de 1943, o médico psiquiatra Warren McCulloch e o matemático Walter Pitts propuseram um modelo matemático simplificado para neurônios artificiais, conforme FIGURA 18, o qual resultou em um sistema com n terminais de entrada $(x_1, x_2, \dots, x_n)$, os quais representavam os dendritos, e o axônio era representado por um sinal de saída (y). (BRAGA, CARVALHO e LUDERMIR, 2011)

FIGURA 18 – NEURONIO DE McCULLOCH E PITTS



FONTE: Adaptado BRAGA, CARVALHO e LUDERMIR (2011).

Para imitar o comportamento das sinapses, os terminais de entrada (x_n) possuíam pesos $(w_1, w_2, \dots, w_n)$ vinculados a cada entrada, podendo estes serem valores positivos ou negativos. O resultado de cada sinapse (i) do neurônio e dado por $(x_i w_i)$, sendo que os pesos determinam o grau de disparo considerado pelo neurônio em determinada conexão. Em um neurônio biológico, a soma dos impulsos que ele recebe determina o grau de disparo, sendo que este disparo ocorre quando esse somatório ultrapassa o limite de excitação (*threshold*) do neurônio. (BRAGA, CARVALHO e LUDERMIR, 2011).

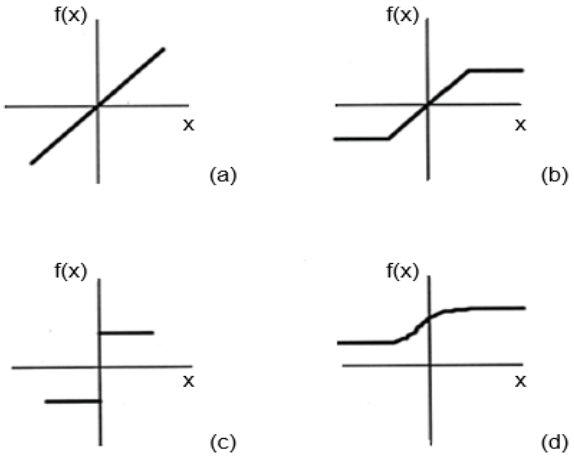
No modelo de neurônios artificiais, a soma ponderada dos valores dos elementos $(x_i w_i)$ recebidos pelo neurônio é o fator de decisão para o disparo, comparando a soma obtida ao limite, ou *threshold* do neurônio, conforme observado na equação (1.20). Esta saída pode variar entre 0 e 1. O processo descrito é considerado como uma função de ativação no qual decide pela soma ponderada das entradas se ativa ou não a saída, onde: n é o número de entradas do neurônio, w_ié o peso associado à entrada x_i e θ é o limiar (*threshold*) do neurônio. (BRAGA, CARVALHO e LUDERMIR, 2011).

$$\sum_{i=1}^{n} x_i w_i \geq \theta \tag{1.20}$$

2.7.2 Funções de ativação

Segundo Braga et. al. (2011), com a evolução das redes neurais artificiais foi possível desenvolver outros métodos, pelos quais eram possíveis uma saída qualquer, não necessariamente 0 e 1, possuindo ainda diferentes funções de ativação das mesmas. Dentre elas é possível destacar a função linear, a função rampa, a função degrau (*step*) e a função sigmoidal, as quais são representadas na FIGURA 19.

FIGURA 19 – FUNÇÕES DE ATIVAÇÃO



FONTE: Adaptado BRAGA, CARVALHO e LUDERMIR (2011).

O item (a) da FIGURA 19 é a função linear, sendo originaria pela equação (1.21), onde y é a saída, x é a entrada e $\propto$ um número real que define uma saída linear para os valores de entrada.

$$y = \propto x \tag{1.21}$$

Quando a função linear é restrita para gerar valores constantes em uma faixa $(-\gamma\ , +\gamma)$ ela passa a se tornar a função rampa como mostra o item (b) da FIGURA 19. Os valores de mínimo e máximo de $-\gamma$ e $+\gamma$ são dados pela equação (1.22):

$$y = \begin{cases} +\gamma & se\ x \geq +\gamma \\ x & se\ |x| < +\gamma \\ -\gamma & se\ x \leq -\gamma \end{cases} \tag{1.22}$$

A função passo, demonstrada no item (c) da FIGURA 19, é semelhante a função sinal levando-se em consideração que ela gera valores positivos de gama para entradas (X) maiores que zero, e valores negativos para gama quando recebem valores de (X) menores do que zero. Esta função é demonstrada na equação (1.23).

$$y = \begin{cases} +\gamma & se\ x > 0 \\ \\ -\gamma & se\ x \leq 0 \end{cases} \tag{1.23}$$

A função sigmoidal, também conhecida como *S-shape* pelo gráfico ter a forma de S, é utilizada para criação de vários modelos e em diversas áreas de aplicação. Esta função é limitada, semi-linear e monotônica, sendo desta forma definir inúmeras funções sigmoidais, sendo uma das mais importantes e aplicadas é denominada de função sigmoidal logística, definida pela equação (1.24):

$$y = \frac{1}{1 + e^{\frac{-x}{T}}} \tag{1.24}$$

Na equação (1.24), o parâmetro T determina a suavidade da curva, sendo o parâmetro de inclinação da função sigmoide. Variando o parâmetro, obtém-se funções sigmoides com diferentes inclinações. (BRAGA, CARVALHO e LUDERMIR, 2011).

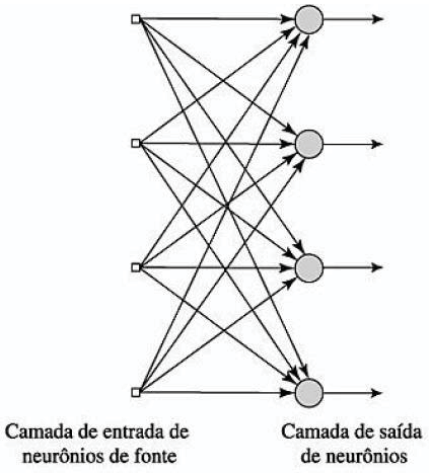
Segundo Haykin (2001), a função sigmoide, é de longe a forma mais comum de função de ativação utilizada na construção de redes neurais artificiais. Ela é definida como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não-linear.

2.7.3 Principais arquiteturas das redes neurais artificiais

Segundo Braga et. al. (2011), baseado no tipo de problema a ser analisado, é importante definir qual a arquitetura de trabalho e a concepção de um RNA, pois a mesma pode restringir o tipo de análise a ser realizada. Fazem parte dos parâmetros de arquitetura de uma rede neural artificial:

- Rede alimentadas com camada única: Nesta estrutura os neurônios estão organizados em forma de camadas, tendo uma forma simples, com apenas uma camada de entrada e uma única camada de saída, porém não ao contrário, sendo essa rede chamada de adiante ou acíclica, conforme demostra a FIGURA 20. (HAYKIN, 2001).

FIGURA 20 – REDE ALIMENTADA ADIANTE OU ACÍCLICA COM UMA ÚNICA CAMADA DE NEURÔNIOS

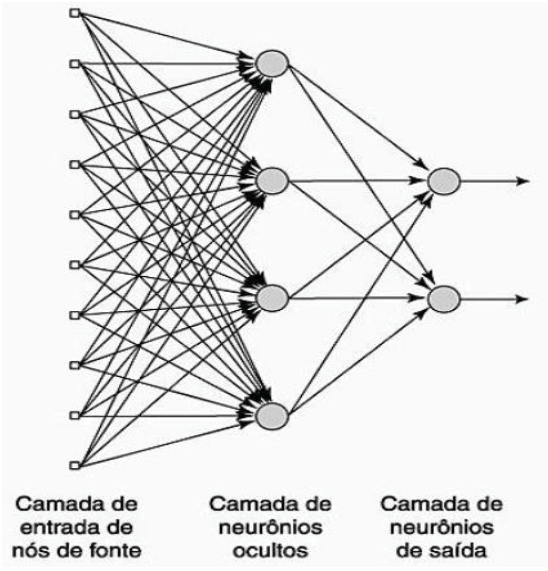


FONTE: HAYKIN (2001).

- Redes alimentadas com múltiplas camadas: Essa estrutura de rede se destaca pela presença de uma ou mais camadas ocultas, conforme demonstrada na FIGURA 21. Um neurônio oculto tem a função de intervir de forma útil entre a camada de entrada e a camada de saída. Essa rede também é considerada como adiante, porém com uma maior quantidade de camadas. Ao incrementar uma maior quantidade de camadas ocultas, os resultados estatísticos se tornam mais elevados. Os nós da camada de

entrada fornecem os padrões de ativação, estes constituem os sinais de entrada da segunda camada, e assim por diante até a última camada e saída da rede. (HAYKIN, 2001).

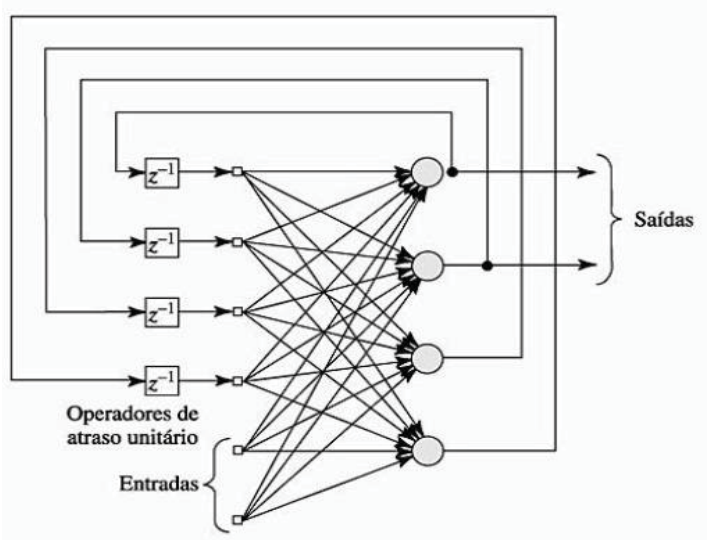
FIGURA 21 – REDE ALIMENTADA ADIANTE OU ACÍCLICA TOTALMENTE CONECTADA COM UMA CAMADA OCULTA E UMA CAMADA DE SAÍDA



FONTE: HAYKIN (2001).

- Redes recorrentes: A rede recorrente se distingue das redes alimentadas adiante por possuir no mínimo um laço de realimentação do sistema, conforme demonstrado na FIGURA 22. Esta rede pode ter apenas uma camada de neurônios, porém tendo o seu sinal de saída sendo enviado para a realimentação dos demais neurônios.

FIGURA 22 – REDE RECORRENTE COM NEURÔNIOS OCULTOS



FONTE: HAYKIN (2001).

2.7.4 Processo de aprendizagem

Uma das propriedades mais importantes das redes neurais é a sua habilidade de aprender e de melhorar a sua performance por meio desse aprendizado. A instrução de uma RNA ocorre com o tempo, seguindo alguns critérios e por meio de elementos de seu ambiente que estimulam o aprendizado da rede, ocorrendo de forma iterativa e com aplicação de pesos em suas sinapses, sendo que após cada processo de iteração, a rede possui um aprendizado cada vez maior. (HAYKIN, 2001).

> Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre. (HAYKIN, 2001, p.75).

O processo de aprendizagem de uma rede neural ocorre inicialmente por uma estimulação do ambiente. Na sequência a rede efetua modificações de seus parâmetros devido aos estímulos recebidos. Por fim, como consequência as modificações realizadas em sua estrutura interna, a rede neural responde ao ambiente de uma nova maneira. (HAYKIN, 2001).

As RNAs possuem a capacidade de apreender por meio de exemplos expostos a elas, e fazer interpolações e extrapolações a partir do que apreenderam. A solução de um problema, ou a análise de uma situação por um modelo de redes neurais inicia-se por uma fase de aprendizagem. Para criar uma representação do problema, a rede identifica e extrai informações relevantes de padrões apresentados a ela. Neste modelo de aprendizado, busca-se identificar a intensidade de conexões entre os neurônios, utilizando um algoritmo de aprendizado. Um algoritmo de aprendizado é um conjunto de procedimentos que possui a finalidade de adaptar os parâmetros de uma rede neural para que esta possa aprender uma determinada função. Para essa atividade, existe uma diversidade de algoritmos, sendo que se diferem basicamente pelos ajustes de pesos utilizados. Cada um dos algoritmos apresenta características para determinada situação exposta, trazendo vantagens ou desvantagens para o modelo. (BRAGA, CARVALHO e LUDERMIR, 2011).

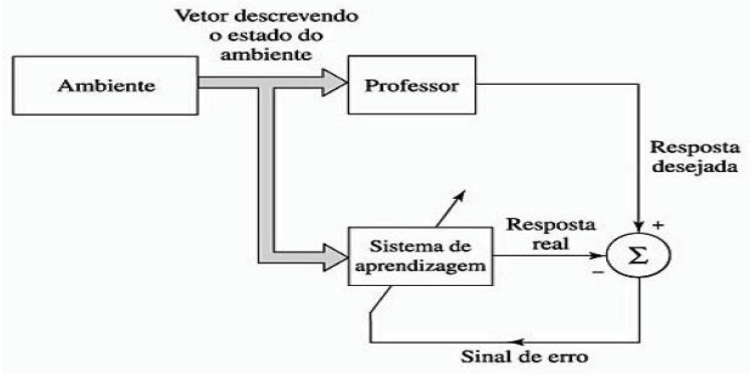
A aprendizagem consiste em um processo pelo qual são realizados ajustes no modelo de RNA de forma contínua e iterativa, podendo ser realizado de forma

supervisionada ou não supervisionada, fazendo com que os parâmetros de peso das conexões entre as unidades de processamento da rede sejam ajustados, para que ao final do processo sejam armazenados os conhecimentos adquiridos pela rede do ambiente em que estava operando. (BRAGA, CARVALHO e LUDERMIR, 2011).

2.7.4.1 Aprendizado supervisionado

O aprendizado de RNA supervisionado consiste em um treinamento pelo qual as entradas e saídas da rede são fornecidas por um supervisor, ou professor, externo, conforme demonstrado na FIGURA 23. Neste método, o supervisor indica comportamentos para a rede, sendo eles tanto bons, quanto ruins, com o objetivo de direcionar o processo de treinamento. Pode ser realizada tanto com pesos, como sem pesos de neurônios, tendo como objetivo encontrar ligações entre os pares de entrada e de saída fornecidos, por meio de ajustes de parâmetros da rede. Como resultado final, a rede possui uma saída calculada comparada com uma saída desejada, gerando assim um erro sobre as respostas. Para cada resultado de erro retornado, comparam-se as respostas desejadas com as respostas calculadas, podendo assim ajustar os pesos do modelo para minimizar os erros finais. Após os ajustes necessários pode ser observado o desempenho da rede, o qual é calculado pela soma dos erros quadráticos de todas as saídas. As desvantagens da utilização das redes supervisionadas estão no fato da dependência de um professor para a rede aprender novas estratégias para situações não expostas nos exemplos de treinamento dos dados. Dentre as técnicas supervisionadas podem ser destacadas a regra delta e a o algoritmo de *backpropagation*, onde é realizada a generalização para redes de múltiplas camadas. (BRAGA, CARVALHO e LUDERMIR, 2011).

FIGURA 23 – DIAGRAMA EM BLOCOS DA APRENDIZAGEM COM UM PROFESSOR

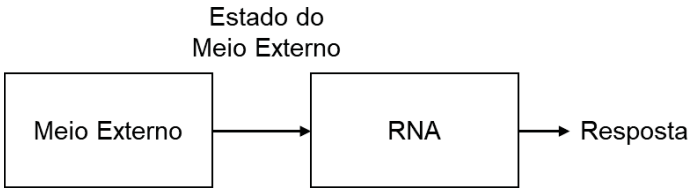


FONTE: HAYKIN (2001).

2.7.4.2 Aprendizado não supervisionado

No aprendizado não supervisionado, ao contrário do supervisionado, não existe um professor ou treinador da rede para acompanhar o processo de aprendizado. Neste modelo, apresentado pela FIGURA 24, apenas os padrões de entrada estão disponíveis para o sistema. No momento em que ocorre uma estabilidade com as regularidades estatísticas de entrada gera-se uma habilidade de desenvolvimento de representações internas sobre as características da entrada da rede, criando novos grupos de forma automática. (BRAGA, CARVALHO e LUDERMIR, 2011).

FIGURA 24 – APRENDIZADO NÃO SUPERVISIONADO



FONTE: Adaptado BRAGA, CARVALHO e LUDERMIR (2011).

2.7.5 Redes MLP – *Perceptron* de Múltiplas Camadas

As redes *perceptron* são os modelos mais simples utilizados em RNAs, sendo compostas basicamente de um único neurônio com pesos ajustáveis e com a função de classificação de padrões linearmente separáveis. (HAYKIN, 2001).

A solução para a resolução de problemas não-lineares requer a utilização de redes com uma ou mais camadas intermediárias, ou ocultas, as quais não constituem as camadas de saída ou entrada da rede. Para treinar as redes que possuem mais de uma camada é utilizado um método que se baseia em gradientes decrescentes, porém para que este funcione de forma eficaz a função de ativação precisa ser contínua, diferençável e de preferência não decrescente. A função de ativação recomendada precisa informar com a maior precisão possível os erros cometidos pela rede pelas camadas anteriores, e para isso, uma alternativa definida é a utilização de funções de ativação do tipo sigmoidal. (BRAGA, CARVALHO e LUDERMIR, 2011).

Segundo Braga et. al. (2011), as redes do tipo *perceptron* multicamadas, ou MLP (*multilayer perceptron*), possuem como características o poder computacional

elevado em relação aos apresentados pelas redes sem camadas intermediárias, e também o fato de tratar dados que não são linearmente separáveis.

Pode ser considerado que as redes MLP funcionam como detectores de características, gerando padrões de entrada que irão resultar em respostas de saída. Para isso é necessário que exista um número suficiente de camadas intermediárias. Em redes com apenas uma camada intermediária, é possível aproximar qualquer função contínua. Em redes com duas, ou mais camadas intermediárias é possível aproximar qualquer função matemática. (BRAGA, CARVALHO e LUDERMIR, 2011).

Usualmente a quantidade de nós das camadas intermediárias são definidos de forma empírica, pois esse número depende diretamente das distribuições dos padrões de treinamento e validação dos dados da rede, pois em alguns casos são necessários apenas uma unidade de entrada e de saída para possuir os padrões requeridos, em outros casos são necessárias diversas unidades intermediárias para isso. Nos casos de problemas práticos de reconhecimento de padrões, adota-se uma quantidade intermediária suficiente para a resolução do problema. (BRAGA, CARVALHO e LUDERMIR, 2011).

Quando se utilizam camadas intermediárias demais pode ocorrer um problema denominado de *overfitting*, o qual se caracteriza pelo elevado número de padrões de treinamento gravados, porém com baixo rendimento no reconhecimento de padrões de saída, pelo excesso de processamento de padrões de treinamento, ao invés do reconhecimento das características de saída da rede. Esse dimensionamento pode dispender um tempo considerável do modelo, pois os números de camadas e de nós são definidos pela complexidade do problema, sendo que a estrutura final é definida por meio de refinamentos sucessivos, uma vez que esse processo é empírico. (BRAGA, CARVALHO e LUDERMIR, 2011).

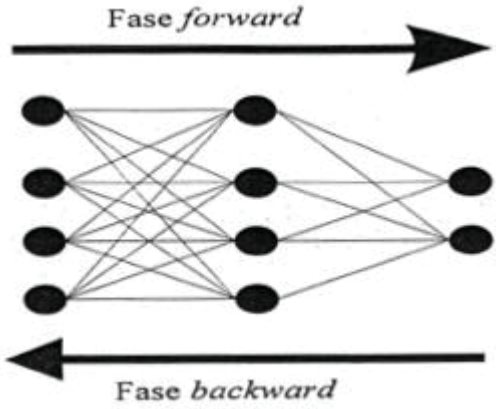
Segundo Braga et. al. (2011), uma forma de evitar o problema de *overfitting* é prever o erro durante o processo de treinamento, definindo conjuntos de dados para treinamento e conjunto de dados para validação do modelo. Dessa forma, o treinamento deve ser interrompido quando a rede iniciar a incorporação dos ruídos presentes nos dados, o que ocasiona a baixa capacidade de validação e previsão dos dados.

Nas redes MLP, o sinal de entrada se propaga para frente, camada por camada da rede, por meio da aplicação de uma rotina de treinamento supervisionada denominada de algoritmo de retropropagação de erro (*erro back-propagation*), sendo

baseado na técnica de aprendizagem por correção de erro, e sendo muito utilizado em problemas considerados como difíceis. (HAYKIN, 2001).

A utilização do algoritmo de retropropagação usualmente constitui-se de dois passos entre as camadas da rede, um sendo para frente (*forward*), denominado de propagação, e outro para trás (*backward*), denominado de retropropagação, conforme demonstrado na FIGURA 25. No passo de propagação, onde os pesos sinápticos são todos fixos, um padrão de entrada percorre a rede camada por camada, sendo gerado um conjunto de saída como resposta real da rede. No segundo passo, o de retropropagação, utilizando a regra de correção de erros, os pesos sinápticos são ajustados, e as respostas reais de saída da rede são subtraídas das respostas desejadas, gerando o sinal de erro, o qual é propagado para trás da rede. Nesse momento, os pesos sinápticos são ajustados para que a resposta real da rede se torne estatisticamente mais próxima da resposta desejada. (HAYKIN, 2001).

FIGURA 25 – FLUXO DE PROCESSAMENTO DO ALGORITMO *BACK-PROPAGATION*



FONTE: BRAGA, CARVALHO e LUDERMIR (2011).

Segundo Braga et. al. (2011), o algoritmo de retropropagação de erro possui etapas em cada uma de suas duas fases:

A fase de propagação consiste nos seguintes passos:

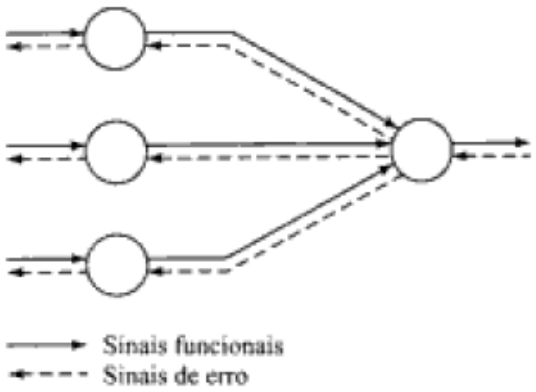
1. Na primeira camada da rede são apresentados os dados de entrada;
2. Após a camada de entrada, os dados de saída de cada camada servem de dados de entrada para a camada posterior;
3. As saídas finais, geradas pela última camada são comparadas com as saídas esperadas.

Na fase de retropropagação, as etapas desde a saída da última camada até o retorno na camada de entrada são executas da seguinte forma:

1. Na camada atual, os nós ajustam seus pesos para reduzir os erros;
2. O cálculo dos erros das camadas intermediárias é realizado pelos erros dos nós da camada seguinte conectados a ele, levando em consideração a ponderação dos pesos das conexões entre eles;

Na FIGURA 26 são demonstradas as direções de dois fluxos de sinais em uma rede *perceptron* de múltiplas camadas, sendo os sinais para frente a propagação, e no sentido contrário a retroprogagação ou os sinais de erro.

FIGURA 26 – FLUXO DE SINAIS EM UMA REDE PERCEPTRON DE MÚLTIPLAS CAMADAS



FONTE: HAYKIN (2001).

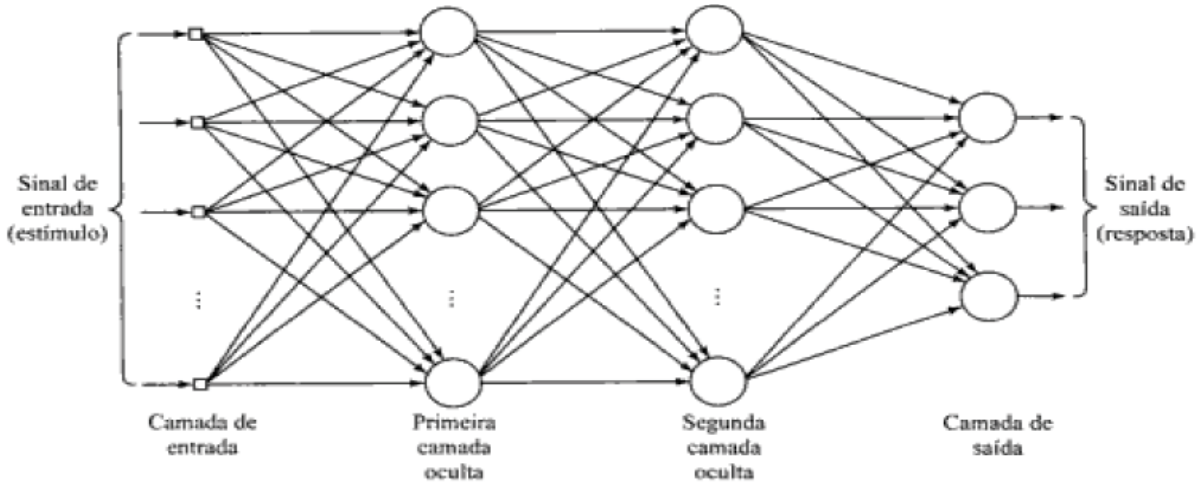
A rede MLP possui três características principais:

1. Diversas funções de ativação têm sido propostas para as redes multicamadas, as quais possuem características de serem não-lineares e diferenciáveis. Tais funções precisam ser diferenciáveis para que os gradientes possam ser calculados e direcionar os ajustes dos pesos. Para as redes MLP, a função de ativação mais utilizada é a sigmoidal logística. (BRAGA, CARVALHO e LUDERMIR, 2011).

2. A rede contém camadas de neurônios ocultos, os quais são os responsáveis por capacitar a rede, treinando-a e extraindo os padrões mais significativos dos dados de entrada. (HAYKIN ,2001).

3. A rede é altamente conectada, fazendo com que qualquer possível alteração na sua conectividade gere mudanças de conexões e dos pesos sinápticos, conforme exemplo da FIGURA 27. (HAYKIN, 2001).

FIGURA 27 – EXEMPLO DE UMA REDE PERCEPTRON COM DUAS CAMADAS OCULTAS



FONTE: HAYKIN (2001).

2.7.6 Aplicação do algoritmo *backpropagation*

Segundo Assef (2018), a utilização do algoritmo de *backpropagation* permite minimizar a função erro, conquistando valores razoáveis de erro. Porém, caso tais valores não sejam aceitáveis, a configuração da rede pode ser alterada, alterando pesos de entrada, ou até introduzindo mais neurônios na rede. Para uma estrutura de duas camadas, o algoritmo de treinamento *backpropagation* possui as seguintes etapas:

1. Fase de Propagação (*Forward*):

    1.1. Etapa de inicialização: determinar valores aleatórios entre [-1,1] para os pesos sinápticos da camada oculta $(W_{kj}\ e\ b_k)$ e para a camada de saída $(W_{lj}\ e\ b_l)$.

    1.2. Etapa de ativação da camada oculta: para cada neurônio k da camada oculta, aplicar a Equação (1.25), onde x_j^p são os dados de entrada da rede.

$$u_k^p(n) = \sum_{j=1}^{m} w_{kj}(n) \times x_j^p(n) + b_k(n) \qquad (1.25)$$

Na sequência é calculado a função de transferência para cada neurônio da camada oculta, conforme Equação (1.26), onde φ usualmente é uma função sigmoide.

$$y_k^p(n) = \varphi_k\left(u_k^p(n)\right) \tag{1.26}$$

1.3. Etapa de ativação da camada de saída: desenvolver a Equação (1.27) para cada neurônio (l) da camada de saída.

$$u_l^p(n) = \sum_{j=1}^{m} w_{lj}(n) \times x_k^p(n) + b_l(n) \tag{1.27}$$

Após, é realizado o cálculo da função de transferência, conforme Equação (1.28), para os neurônios da camada de saída, onde φ usualmente é uma função linear.

$$y_l^p(n) = \varphi_l\left(u_l^p(n)\right) \tag{1.28}$$

2. Fase de retropropagação (*backward*):

2.1. Etapa de ajuste dos pesos da camada de saída para a camada oculta: pela Equação (1.29) calcula-se o erro da camada de saída, onde (d_j) é o valor previsto de saída e (φ') é a derivada da função de transferência do neurônio de saída.

$$\delta_l^p(n) = \left(d_j^p - y_l^p\right) \times \delta_l'\left(u_l^p(n)\right) \tag{1.29}$$

Na sequência é desenvolvida a Equação (1.30), a qual fornece o ajuste a ser realizado no peso.

$$\Delta_{w_{lj}}^p = \beta\left(w_{lj}(n-1)\right) + \eta\delta_l(n)y_l(n) \tag{1.30}$$

Após ter os valores de ajuste, os pesos sinápticos da camada de saída devem ser ajustados, conforme Equação (1.31), onde (n) é a taxa de aprendizagem e (β) a taxa de *momentum*.

$$w_{lj}(n+1) = w_{lj}(n) + \Delta^{p}_{w_{lj}} \tag{1.31}$$

2.2. Etapa de ajuste dos pesos da camada oculta para a camada de entrada: neste passo é calculado o erro da camada oculta por meio da Equação (1.32).

$$\delta^{p}_{k}(n) = \varphi'_{k}\left(y^{p}_{k}(n)\right)\sum_{l}\delta^{p}_{l}(n)w_{lj}(n) \tag{1.32}$$

De forma semelhante a etapa 2.1, calcular a Equação (1.33) para obter os pesos a serem atualizados na Equação (1.34).

$$\Delta^{p}_{w_{kj}} = \beta\left(w_{kj}(n-1)\right) + \eta\delta_{k}(n)y_{k}(n) \tag{1.33}$$

$$w_{kj}(n+1) = w_{kj}(n) + \Delta^{p}_{w_{kj}} \tag{1.34}$$

3. Fase de cálculo do erro global: Após ser realizada a primeira iteração, ou seja, todos os exemplos de treinamento terem passado pela rede, é realizado o cálculo do erro global conforme Equação (1.35).

$$\varepsilon(n) = \frac{1}{N}\sum_{l=1}^{N} e^{2}_{l}(n) \tag{1.35}$$

2.7.7 Critérios de parada

Segundo Haykin (2001), não existe uma forma visual ou prática para demonstrar em que momento o algoritmo de retropropagação pode ser parado. Porém existem alguns critérios de parada que podem ser adotados para encerrar o ajuste de pesos do modelo, podendo assim utilizar o seguinte critério de convergência:

> Considera-se que o algoritmo de retropropagação tenha convergido quando a taxa absoluta de variação do erro médio quadrado por época for suficientemente pequena. (HAYKIN, 2001, p.200).

Uma taxa de erro médio pode ser considerada suficientemente pequena se ela estiver no intervalo entre 0,1 e 1 por cento. (HAYKIN, 2001).

Segundo Braga et. al. (2011), existem várias técnicas para dizer em que momento o treinamento deves ser interrompido, sendo os mais utilizados:

- Após um determinado número de iterações realizadas;
- No momento em que o erro quadrático médio ficar abaixo de uma constante determinada;
- Combinação dos métodos supracitados.

Segundo Braga et. al. (2011), na aplicação das redes neurais artificiais, as redes *perceptron* multicamadas (MLP) são as mais utilizadas pela sua simplicidade e facilidade de implementação, sendo evidenciadas sua utilização em aplicações como: reconhecimento de caracteres, previsões de comportamentos, verificação de assinaturas, segurança de transações, diagnósticos variados, entre outras.

3 MÉTODO DE PESQUISA

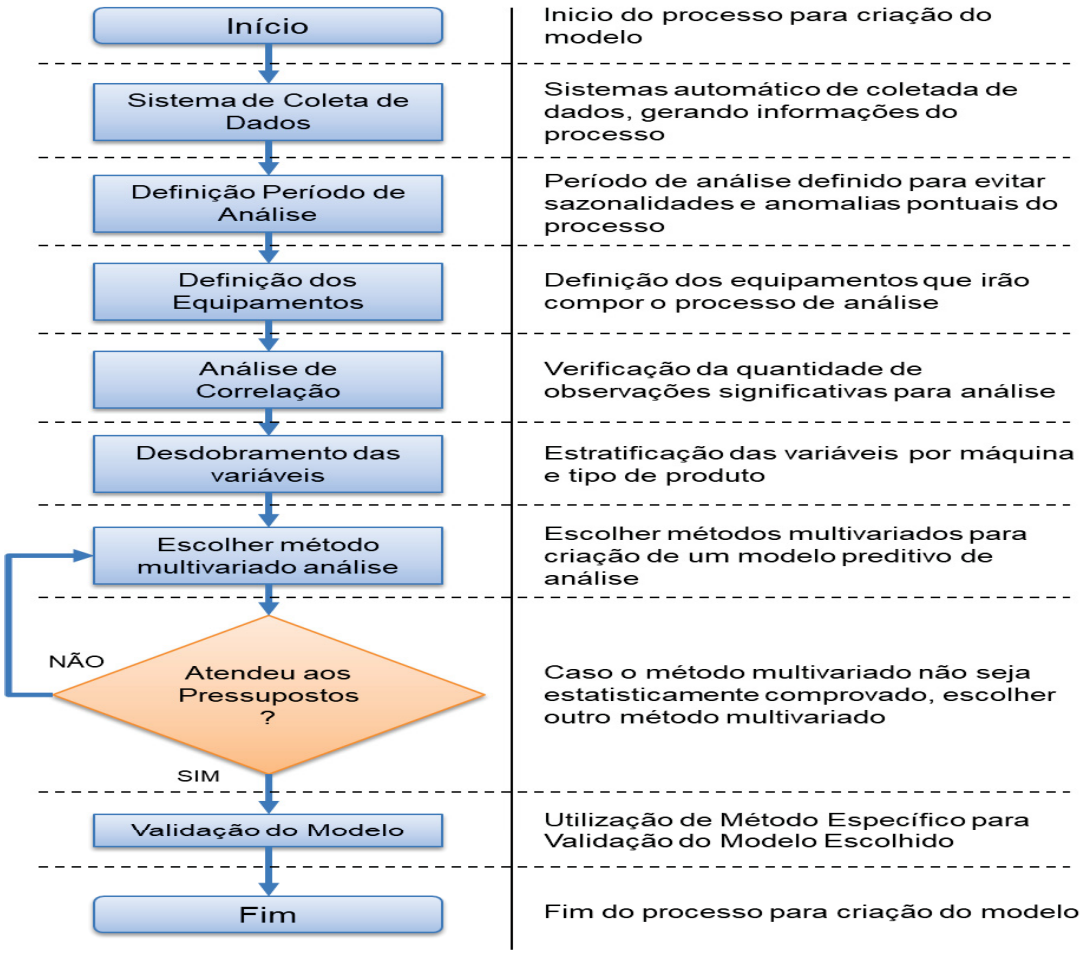
Este capítulo tem por objetivo apresentar e descrever as metodologias aplicadas neste trabalho, sendo expostos o sistema para a coleta, preparação e análise dos dados, além da apresentação das formas de atuação das técnicas estatísticas e multivariadas empregadas.

3.1 PROTOCOLO DE COLETA E ANÁLISE DOS DADOS

Nesta seção serão apresentadas as etapas metodológicas para a criação de um modelo preditivo de OEE, sendo divididas em sistema para coleta de dados, amostra, e sistema de análise dos dados para a elaboração do modelo.

O fluxo demonstrado na FIGURA 28 descreve o planejamento para a elaboração do modelo preditivo. Cada uma das etapas deste fluxo será apresentada nas próximas seções.

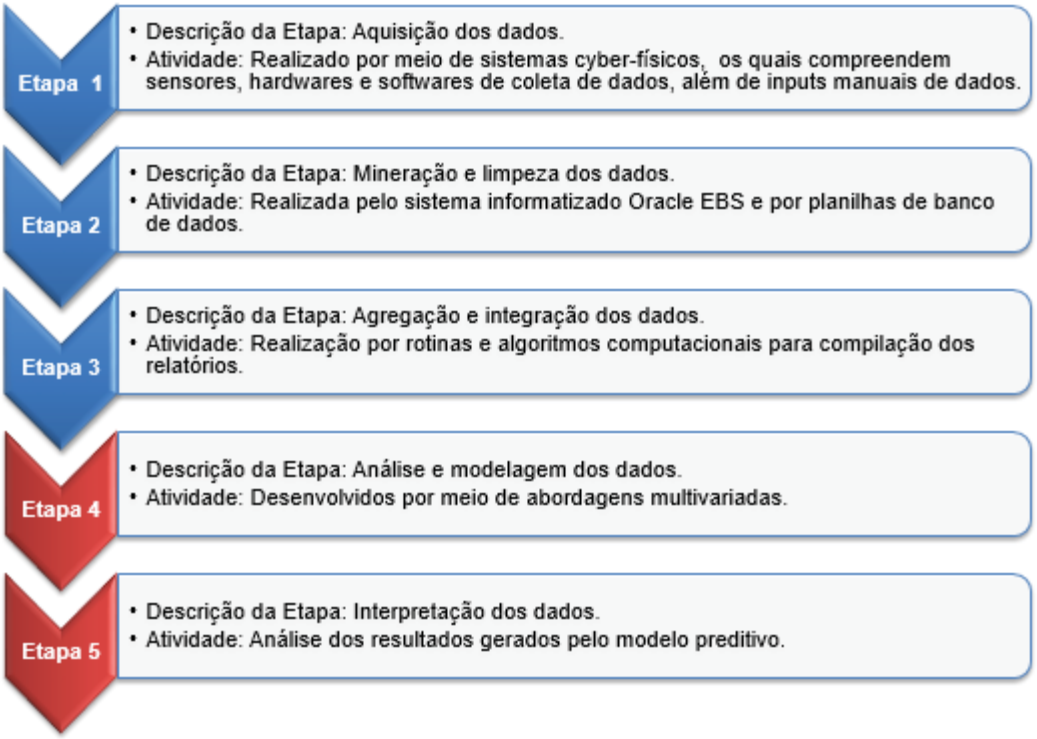
FIGURA 28 – FLUXOGRAMA PARA ELABORAÇÃO DO MODELO PREDITIVO



FONTE: O autor (2019).

Nas fases referentes ao desenvolvimento do modelo, será utilizado como referência as etapas descritas por Sivarajah et al. (2016), as quais definem que o processo de análise de *Big Data* é dividido em cinco etapas principais. Destas etapas, conforme demonstrado na FIGURA 29, o presente trabalho irá focar de forma mais incisiva nas fases quatro e cinco do processo.

FIGURA 29 – ETAPAS PARA ANÁLISE DOS DADOS



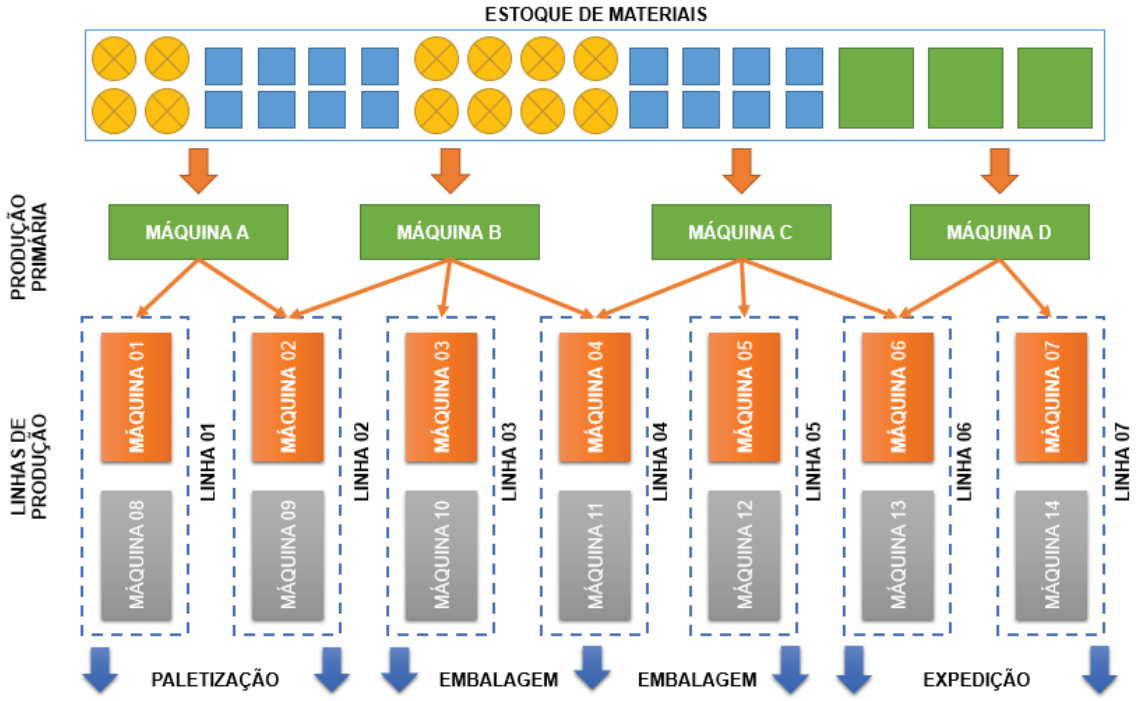
FONTE: Adaptado de SIVARAJAH et. al. (2016).

A primeira etapa é realizada no ambiente fabril, por meio de sistemas *cyber-físicos* compostos por sensores, interfaces homem-máquina, *hardwares*, *softwares*, entre outros, além da interação humana em alguns apontamentos complementares de produção. Na segunda etapa, após os dados serem inseridos no sistema informatizado, a própria plataforma de banco de dados realiza a mineração e limpeza destes dados para posteriormente na terceira fase, por meio de rotinas e algoritmos realizar-se a agregação e integração destas informações ao banco de dados, os quais disponibilizarão relatórios gerenciais que possuem as informações necessárias para a realização das etapas posteriores. Na quarta etapa serão utilizadas técnicas multivariadas para o tratamento dos dados e na quinta etapa será realizada as interpretações dos resultados de saída do modelo preditivo.

3.1.1 Coleta de dados

O presente estudo utilizou-se de informações de produção oriundas do processo de fabricação de embalagens de papel. A fábrica é composta por sete linhas de fabricação, sendo que nessas linhas existem sub-processos de fabricação independentes, conforme demonstrado no diagrama ilustrado na FIGURA 30.

FIGURA 30 – FLUXO PRODUTIVO

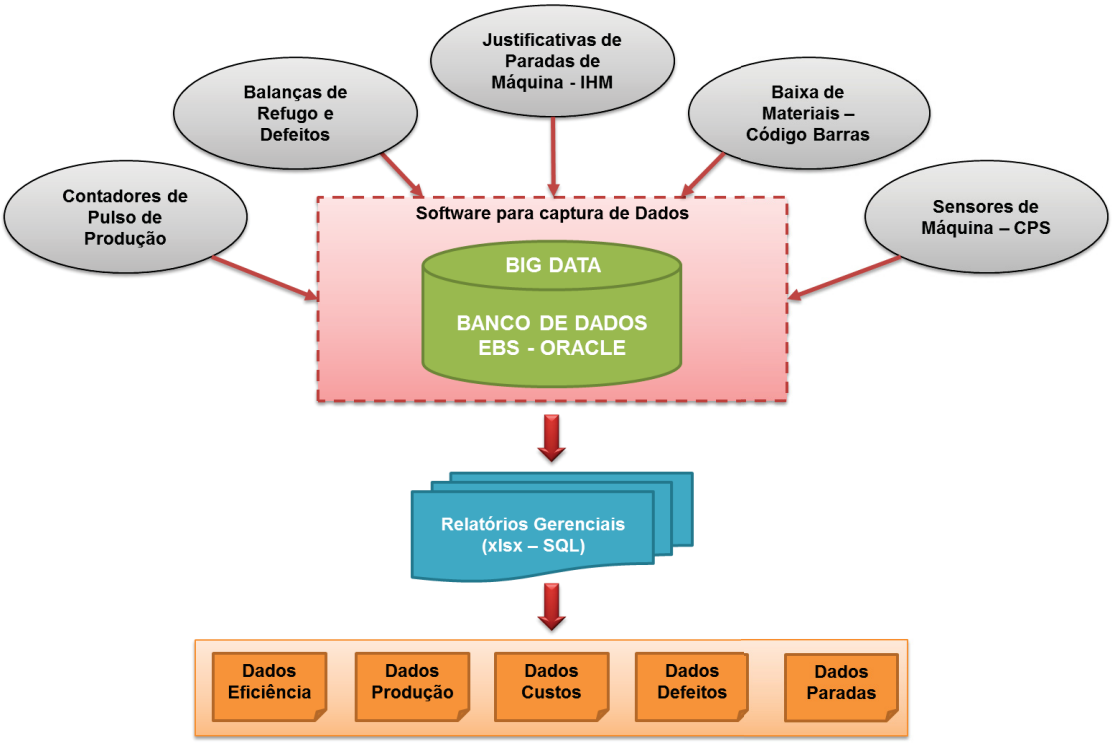


FONTE: O autor (2019).

Na produção primária, ocorre o processo de sub-montagem do produto, sendo na sequência enviado para a próxima fase (as linhas de produção), dependendo do roteiro estipulado pela engenharia de produto e planejado pelo PCP (Planejamento e Controle de Produção). Cada um dos equipamentos possui sistemas de coletas de dados de produção independentes, os quais geram informações sobre os equipamentos e sobre o processo produtivo.

A coleta de dados ocorre por meio de um compilado de informações que são originadas em sua grande maioria de forma automática por meio de sistemas *cyber-físicos*, e que posteriormente são enviadas para um banco de dados por meio de *softwares* específicos para esse fim. O fluxo para a forma de coleta de dados é apresentado na FIGURA 31.

FIGURA 31 – FLUXO PARA COLETA DE DADOS



FONTE: O autor (2019).

Pelo conceito da Internet das Coisas e do *Cyber-Physical System* (CPS) os pulsos eletrônicos gerados pelos equipamentos são enviados por meio de redes industriais para sistemas onde são transformados em dados com o auxílio de *softwares* especializados. Tais dados são armazenados em um banco de dados em uma plataforma *Oracle*.

As informações relevantes para a pesquisa foram extraídas do banco de dados do sistema, onde é possível gerar relatórios gerenciais, os quais possuem informações cronológicas sobre todos os eventos ocorridos na fábrica durante cada fase do processo.

3.1.2 Caracterização da amostra

Para o processo de pesquisa foram coletados dados referentes de janeiro de 2015 a dezembro de 2017, representando três anos de informações.

As informações exportadas do banco de dados possibilitam gerar relatórios gerenciais a respeito de todos os itens fabricados em um determinado período de

tempo. Nestes relatórios existem variáveis amostrais que são constituídas por características de produto e processo, compostas por:

- Descrição e código do produto;
- Ordem de produção;
- Estrutura, formato e configuração do produto;
- Recurso empregado;
- Equipamento utilizado na produção;
- Sequenciamento de produção;
- Equipe e turno de trabalho;
- Tempo de *setup*;
- Produção realizada;
- Quantidade de refugos e defeitos gerados;
- Paradas de máquina programadas e não programadas;
- Motivo das paradas de máquina;
- Velocidade padrão e realizada por produto.

Estas informações estão distribuídas em dois relatórios de dados, um referente aos dados de produção e refugo, e o outro sobre dados de paradas dos equipamentos. Para o estudo foram selecionados os dados referentes a três anos de produção por possuir um período histórico considerável de dados de fabricação, além de evitar interferências de sazonalidade e de possíveis anomalias ocorridas neste período.

Os dados contidos no relatório de produção e refugo estão distribuídos em 37 variáveis (colunas), as quais possuem em média cerca de 28.500 observações geradas por mês, totalizando uma quantidade média mensal de mais de 1 milhão de dados. O segundo relatório, destinado a informações de paradas de máquina, possui 32 variáveis, com uma média de 56,7 mil observações geradas mensalmente, totalizando uma quantidade média de 1,8 milhões de dados. A magnitude média fornecida de informações por estes relatórios está na faixa de 2,9 milhões de dados mensais, o que equivale a quase 34,4 milhões de dados por ano e a aproximadamente 103,2 milhões de dados no período de três anos.

Tais dados referem-se a todos os apontamentos de produção gerados pelos 17 principais equipamentos contidos no parque fabril, porém tais equipamentos possuem características mecânicas construtivas bem distintas, as quais poderiam ser um empecilho para a criação do modelo, pois não são comparáveis para a geração

de um padrão. Tendo tal definição, escolheram-se os 7 equipamentos da saída do processo, os quais geram os produtos acabados. Desta forma foi possível comparar os dados destes equipamentos, pois possuem características semelhantes, além de que considerando máquinas do fim do fluxo de processo, foi possível avaliar os dados de produtos completos, ou seja, produtos prontos para serem embalados e expedidos.

Após a definição da base de dados a ser trabalhada, tornou-se necessário a compilação de todos os relatórios gerados do período escolhido. Nesse momento foi necessário realizar uma limpeza dos dados, retirando os equipamentos que não faziam parte da amostra selecionada. Para tal atividade foi utilizado o *software Excel*, onde se utilizou rotinas em *Visual Basic* (VBA) para eliminar as observações desnecessárias.

Nos relatórios de dados também existiam variáveis que não possuíam influência no OEE dos equipamentos. Inicialmente, contabilizando os dois relatórios, existiam 69 variáveis de dados. Destas variáveis, 23 foram selecionadas como possíveis preditoras do OEE. Tais variáveis foram selecionadas por possuírem potencial interferência na eficiência do equipamento, conforme demonstrado no QUADRO 1.

QUADRO 1 – VARIÁVEIS PREDITORAS PRELIMINARES

(continua)

Variável	Descrição	Código	Unidade
Produção	Taxa de produção realizada em um período de tempo	Prod	Unidade (x10.000)
Velocidade Real	Velocidade realizada na produção do item	V_Real	Unidades / Minuto
Velocidade teórica	Velocidade planejada para a fabricação do item	Vel_T	Unidades / Minuto
Horas de Produção	Tempo realizado para fabricação do item	H_Prd	Horas / Evento (x10)
Horas de *setup*	Tempo realizado para a troca de pedido	H_Stp	Horas / Evento
Paradas Operacionais	Intervenções não planejadas ocorridas nos equipamentos pela equipe de produção, como: ajustes, limpeza, trocas frequenciadas, etc	Par_Op.	Horas / Evento
Paradas Não Programadas	Intervenções não planejadas ocorridas nos equipamentos pelas equipes de apoio a produção, como por exemplo, quebras de equipamento, falta de abastecimento, etc	Par_N.Prog.	Horas / Evento

QUADRO 1 – VARIÁVEIS PREDITORAS PRELIMINARES

(conclusão)

Variável	Descrição	Código	Unidade
Paradas Programadas	Paradas planejadas pelo PCP para os equipamentos, como por exemplo, refeição, manutenções preventivas planejadas, etc	Par_Prog.	Horas / Evento
Defeitos	Refugos de produção gerados no processo ou no *setup* do equipamento	Def	Unidades (x100)
Máquina	Recurso empregado para a fabricação do produto	Maq	Decimal
OEE da Máquina	OEE global do equipamento considerando todos os produtos produzidos na máquina em determinado período	OEE_Maq	Decimal
Mês	Intervalo de tempo em que o produto foi fabricado	Mês	Mês
Turno	Turno de trabalho em que o produto foi fabricado	Turno	Turno
Peso	Peso do produto fabricado, considerando os materiais e os insumos utilizados	Peso	gramas
Largura Fechado	Largura do produto acabado	Larg_F	cm
Altura Fechado	Altura do produto acabado	Alt_Fec	cm
Largura Aberto	Largura do produto no início de fabricação	Larg_A	cm
Altura Aberto	Altura do produto no início de fabricação	Alt_A	cm
Gramatura	Gramas de papel por centímetro quadrado	Gram	g/cm^2
Tipo de Corte	Característica estrutural do produto, sendo este o modelo de corte para fechamento da embalagem	T_Corte	cm
Tipo de Fundo	Característica estrutural do produto, sendo este o modelo do fechamento do fundo da embalagem	T_Fundo	cm
Tipo de Válvula	Característica estrutural do produto, sendo este o modelo de Válvula para enchimento da embalagem	T_Valv	cm
Escalonamento	Característica estrutural do produto, sendo este o tipo de sobreposição para fechamento da embalagem	Esc	cm

FONTE: O autor (2019).

Tais variáveis foram escolhidas devido as possíveis influências destas no comportamento do OEE. Estas 23 variáveis podem ser utilizadas para a predição do OEE conforme o desdobramento desejado, podendo ser dividido por máquina, período, turno, estrutura do produto, entre outros. Vale ressaltar que se a análise ocorrer sobre um determinado tipo de produto, a estrutura do produto se torna irrelevante, pois os dados das observações se tornam constantes.

3.1.3 Passos para aplicação dos métodos

O processo de para a construção do modelo preditivo consiste em etapas de análises estatísticas ou multivariadas para buscar uma técnica que seja adequada aos pressupostos estatísticos e que retorne resultados com índices aceitáveis de erro. Desta forma trabalhou-se com algumas abordagens até se chegar a um modelo apropriado para previsão e análise do comportamento do OEE em relação às variáveis preditoras. Quando algum pressuposto estatístico não era atendido, ou quando um erro de saída retornava valores altos, buscava-se outra técnica ou outra configuração para o desenvolvimento do modelo, as quais são descritas nas seções seguintes.

3.1.3.1 Análise de correlação

Uma das questões fundamentais para a análise dos relatórios, e para um melhor e mais rápido processamento dos dados, é a definição de qual a quantidade mínima de observações necessária para se realizar uma análise estatisticamente significativa, caso fossem utilizadas todas as observações disponíveis. Para responder a esse questionamento utilizou-se como base a matriz de correlação.

Para o cálculo da matriz de correlação utilizou-se o *software* R. Como base para este processo utilizou-se o p-valor, para confirmar se existem correlações significativas entre os pares de variáveis. Nas análises considerou-se um nível de significância de 5% para verificação da existência de correlação significativa.

Inseriu-se os dados em um algoritmo no *software* R (APÊNDICE A), sendo realizadas simulações para as quantidades de 100, 500, 1000, 5000, 10000, 25000, 50000, 100000, 250000, 500000 e para o valor total de observações, com o intuito de verificar o comportamento do p-valor conforme as observações eram incrementadas. Nessas análises, também se levou em consideração a aleatoriedade dos dados,

sendo que as análises ocorreram com os dados em ordem cronológica e também de forma aleatória.

Tendo a percepção de que os p-valor*es* iriam variar conforme aumentavam-se as observações de dados, gerou-se um algoritmo no *software* R para observar esse comportamento de uma forma mais contínua e com pequenos incrementos ao logo de toda a amostra. Para isso, o algoritmo calculou os p-valores em incrementos de 100 observações até chegar à quantidade total de observações.

Para se verificar o comportamento do p-valor, realizou-se o mesmo processo, com as observações retiradas de três dos sete equipamentos escolhidos como amostra. Os três equipamentos selecionados, M11, M12 e M14, foram escolhidos por possuirem uma maior diversidade de produtos produzidos, reduzindo assim tendências em relação ao mix de produção.

3.1.3.2 Escolha dos métodos estatísticos e multivariados – Modelo I

Antes de se utilizar qualquer análise estatística por técnicas de dependência, buscou-se por meio de análise da matriz de correlações possíveis variáveis que não possuíam correlação estatisticamente significativa. Localizando-se tais variáveis, as mesmas poderiam ser eliminadas do modelo por terem pouca significância estatística para a análise.

Outro retorno esperado pela análise da matriz de correlação é em relação a quantidade de observações relevantes para o modelo. Espera-se que com a simulação do cálculo da matriz de correlação para diferentes quantidades de observações possa existir uma quantidade de observações mínima, a qual não prejudique o resultado final do modelo, mas que também traga benefícios tais como, ganho computacional e menor tempo no processo de análise.

Por meio dos resultados da análise da matriz de correlação, o processo inicial para a construção do modelo foi à escolha dos métodos estatísticos e multivariados para auxiliar na construção do modelo preditivo. Inicialmente foram escolhidos os métodos de análise de componentes principais com o objetivo de eliminar variáveis não representativas para o modelo, e também a técnica de regressão linear múltipla, com o objetivo de determinar uma equação preditiva baseada nos dados amostrais.

3.1.3.2.1 Análise de componentes principais – Modelo I

O propósito da utilização da técnica de componentes principais em conjunto com o método de Jolliffe B2 e B4 é a redução do número de variáveis do conjunto de dados. Para o cálculo das componentes principais, foi necessário gerar os autovalores e autovetores das matrizes de correlações.

Nesta etapa foi utilizando o *software* R e um algoritmo para realizar a análise de componentes principais. Tal algoritmo retorna os autovalores das matrizes de correlações, as componentes principais (autovetores), além do valor do desvio padrão e das variâncias encontradas para cada componente.

Após a seleção das variáveis por componentes principais em conjunto com o método de Jolliffe B2 e B4, o próximo passo é realizar a análise de regressão linear múltipla.

O objetivo da análise de regressão linear múltipla é gerar uma equação preditiva para o OEE, porém não é possível gerar uma equação única considerando todos os equipamentos, pois apesar de semelhantes, as máquinas possuem fatores específicos em relação a sua construção mecânica, com fatores externos ao processo, como interferências da mão-de-obra, métodos de trabalho não padronizados ou cumpridos, sazonalidades de produção, mix de produtos diferenciados, entre outros, os quais são inerentes ao processo.

Em virtude dessa reflexão, e visando possuir equações mais realistas a respeito dos equipamentos, realizou-se uma análise sobre a quantidade de observações em relação aos equipamentos, aos turnos de trabalho e ao período de fabricação, para escolher um conjunto de equipamentos distintos, gerando uma equação única para cada equipamento. Como o OEE reflete a eficiência de um determinado equipamento, tal decisão vai de encontro aos conceitos da eficiência global do equipamento.

3.1.3.3 Escolha dos métodos estatísticos e multivariados – Modelo II

Definido os equipamentos, antes de se realizar a análise de regressão, optou-se por realizar novamente as etapas de análise da matriz de correlação e análise de componentes principais, uma vez que foram retiradas observações da base de dados referente aos equipamentos que não seriam analisados.

3.1.3.3.1 Análise de regressão linear múltipla – Modelo II

Após a aplicação da técnica de componentes principais e Jolliffe B2 e B4, definiram-se quais variáveis permaneceriam na análise de regressão. Após essa etapa, efetuou-se a análise de regressão linear múltipla nas variáveis selecionadas, sendo esta realizada em cada um dos equipamentos selecionados. Para essa etapa foi utilizado novamente o *software* R, o qual retornou os coeficientes estimados de cada uma das variáveis preditoras, os p-valores de cada um dos coeficientes estimados e o coeficiente de determinação (R^2), o qual representa o ajuste de um modelo estatístico linear em comparação ao conjunto de dados analisados.

Neste cenário, o próximo passo foi efetuar a análise dos resíduos do modelo estimado por meio dos testes de normalidade, homocedasticidade e independência dos resíduos.

3.1.3.4 Escolha dos métodos estatísticos e multivariados – Modelo III

Segundo Stamatis (2010), uma das características do OEE refere-se ao seu padrão de distribuição dos dados de forma não-linear, desta forma buscou-se por outro método multivariado que apresentasse por característica a análise de dados com padrões não-lineares e que não se encaixassem em uma distribuição normal. Nesta situação optou-se pela utilização das redes neurais artificiais.

3.1.3.4.1 Escolha dos métodos estatísticos e multivariados - RNA

Para a utilização das redes neurais artificiais com o objetivo de construir um modelo preditivo para a análise de OEE, seguiu-se com a mesma linha de raciocínio, porém sem perder ou retroceder com o aprendizado adquirido nas duas tentativas anteriores. Para a análise da RNA, optou-se por realizar em um único equipamento, pois a solução apresentada neste poderia ser replicada para os demais. Seguindo as mesmas lógicas apresentadas nos modelos I e II, optou-se por desenvolver o modelo por RNA no equipamento M14. Desta forma, a análise focou em uma família de produtos deste equipamento, sendo que por essa abordagem possuíamos variáveis constantes em relação a estrutura do produto, reduzindo assim as variáveis preditoras para o modelo. De tal forma, restaram dez variáveis para a elaboração do modelo, as quais são descritas no QUADRO 2.

QUADRO 2 – VARIÁVEIS PREDITORAS UTILIZADAS NA RNA

Variável	Descrição	Código	Unidade
Produção	Taxa de produção realizada em um período de tempo	Prod	Unidade (x10.000)
Velocidade Real	Velocidade realizada na produção do item	V_Real	Unidades / Minuto
Horas de Produção	Tempo de produção realizado para cada lote ou apontamento de produção	H_Prod	Horas / Evento (x10)
Horas de *setup*	Tempo realizado para a troca de pedido	H_Stp	Horas / Evento
Paradas Operacionais	Intervenções não planejadas ocorridas nos equipamentos pela equipe de produção, como por exemplo, ajustes, limpeza, trocas frequenciadas, etc	Par_Op.	Horas / Evento
Paradas Não Programadas	Intervenções não planejadas ocorridas nos equipamentos pelas equipes de apoio a produção, como por exemplo, quebras de equipamento, falta de abastecimento, etc	Par_N.Prog.	Horas / Evento
Paradas Programadas	Paradas planejadas pelo PCP para os equipamentos, como por exemplo, refeição, manutenções preventivas planejadas, etc	Par_Prog.	Horas / Evento
Defeitos	Refugos de produção gerados no processo ou no *setup* do equipamento.	Def	Unidades (x100)
Velocidade Teórica	Velocidade prevista na produção do item	V_Teo	Unidades / Minuto
OEE da Máquina	OEE global do equipamento considerando todos os produtos produzidos na máquina em determinado período	OEE_Maq	Decimal

FONTE: O autor (2019).

Como o modelo pode ser contínuo para cada família de produto, o seu passo a passo também pode ser replicado para as demais famílias, sem perder as suas características construtivas.

Novamente foi utilizado o *software* R em um computador *workstation* Xeon ES 1650 V4, 3.60 GHz, 64 GB de RAM, mais especificamente o pacote *neuralnet*, para a aplicação das redes neurais, onde se utilizou a rede do tipo MLP (*MultiLayer Perceptron*), com uma camada oculta e aplicando sigmoide logística como função de ativação na camada oculta e função linear na camada de saída. Também foi utilizado

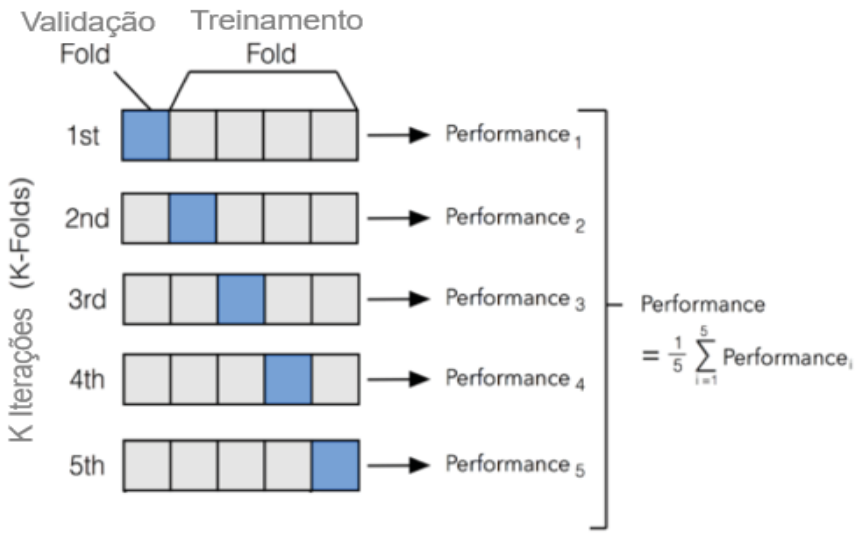
0,01 como *threshold* do neurônio e número máximo de iterações como 100.000. Todos esses parâmetros são *default* da função *neuralnet* do R.

Com a disponibilidade de dez variáveis preditoras para o equipamento M14, a camada de entrada possuía 10 neurônios, referentes as dez variáveis. A camada de saída possuía apenas um neurônio, referente ao OEE previsto. Com as dez variáveis de entrada, busca-se modelar e predizer o comportamento da variável de saída (OEE).

Segundo Braga et. al. (2011), o número de neurônios da camada oculta pode ser escolhido de forma empírica, desta forma variou-se esse parâmetro de 1 a 20, sendo realizadas 15 simulações para cada uma das redes. Fez-se a separação entre conjunto de dados para treinamento e testes da rede de forma aleatória, os quais foram considerados 20% dos dados para a fase de testes e 80% dos dados para a etapa de treinamento. Seguindo essa lógica, na rede com um neurônio na camada oculta foram aplicadas 15 simulações. Na rede com dois neurônios na camada oculta foram aplicadas outras 15 simulações, e assim por diante até o número de neurônios na camada oculta chegar a 20. Feitas as $15 \times 20 = 300$ simulações, calculou-se o erro de previsão para o conjunto de teste, assim por exemplo, para as redes com um neurônio na camada oculta foram calculados 15 erros de previsão para então se calcular a média dos erros. Portanto, para as 20 configurações das redes, tem-se um erro médio de previsão. A partir desses erros, calculou-se o menor dentre eles para a escolha do melhor número de neurônios na camada oculta.

Finalizada a escolha do número de neurônios na camada oculta, foram realizadas as etapas de validação das redes. Nesta fase, para cada item da família de produtos selecionada, foi aplicada a técnica conhecida como *k-fold* (KOHAVI, 1995). Esta técnica consiste em dividir o conjunto total de dados em k subconjuntos disjuntos de mesmo tamanho, onde $k-1$ subconjuntos são usados para treinamento e o conjunto restante é usado para teste. O processo é realizado k vezes alternando o subconjunto de teste. A FIGURA 32 ilustra no processo *k-fold*.

FIGURA 32 – ESQUEMA DE VALIDAÇÃO *K-FOLD*



FONTE: Adaptado RASCHKA (2016).

Assim, depois de embaralhar os dados e usando $k = 5$, 80% destes foram usados para o treinamento (já com um número de neurônios na camada oculta escolhido na fase anterior) e 20% para teste e cálculo do erro. Depois disso, outros 80% dos dados foram usados para o treinamento e outros 20% para o cálculo do erro, e assim por diante até que se obteve 5 valores de erro para cada rede especializada em um produto específico da família de produtos. Neste trabalho, considerou-se como uma rede neural artificial validada quando resultasse em um erro de previsão médio abaixo dos 3%.

Para verificar a influência de cada uma das variáveis preditoras na previsão do OEE, fixou-se nove das dez variáveis de entrada pela mediana de cada variável, e criou-se uma sequência de pontos para a variável não fixada, iniciando-se pelo seu valor mínimo, até a finalização em seu valor máximo. Desta forma, essas 10 variáveis, sendo uma fixada e as demais variando, formaram os dados de entrada para cada uma das redes das famílias dos produtos A e produto B, resultando assim em uma previsão de valores para o OEE baseado em cada uma das variáveis preditoras.

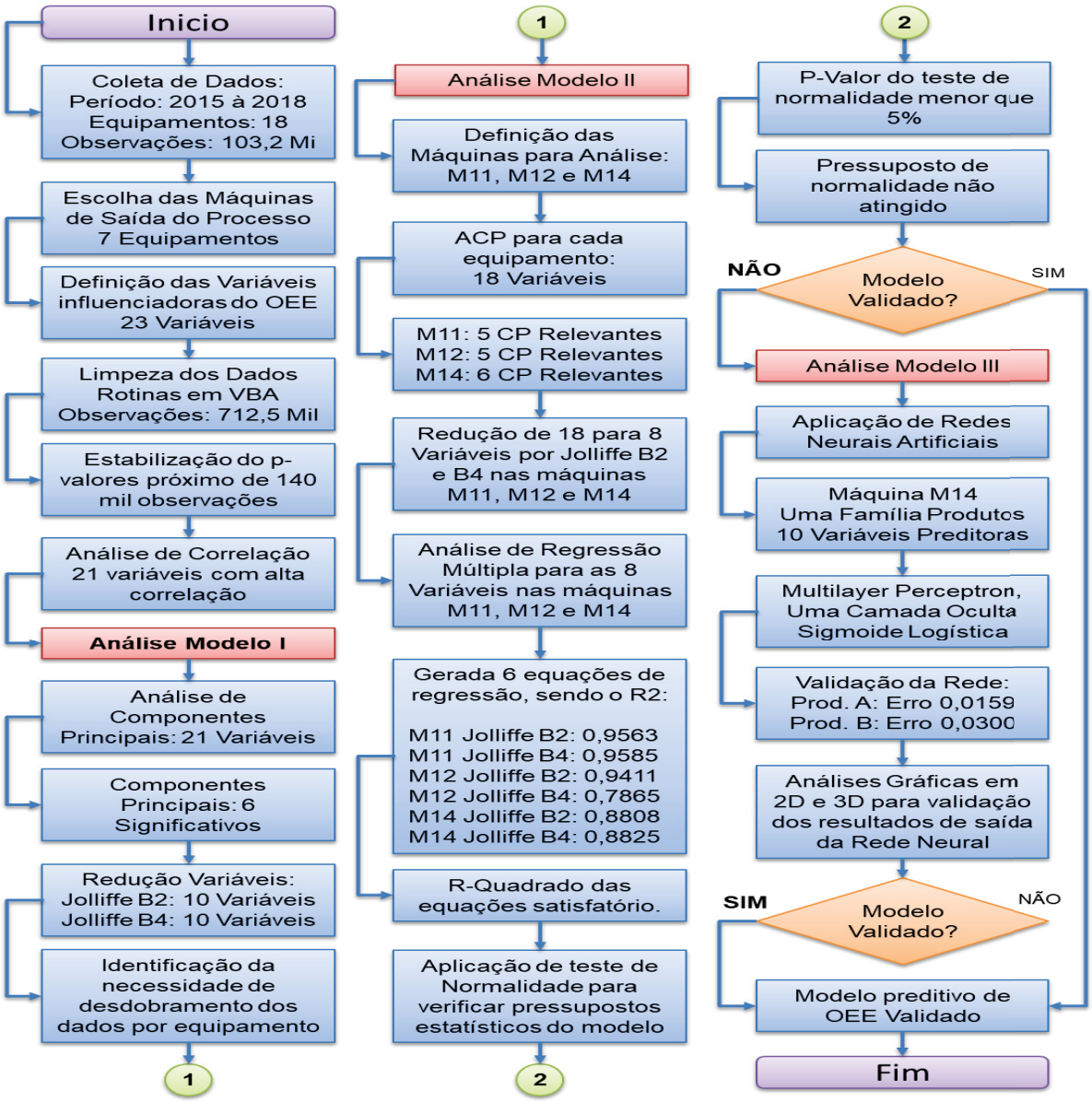
Para explicar melhor alguns eventos de previsão do OEE foram utilizados duas variáveis fixas e as demais variando. Dessa forma foi possível criar gráficos em 3D para um melhor entendimento dos fenômenos que poderiam influência de alguma forma a previsão do OEE para a família de produtos. Tais gráficos foram criados com o auxílio do *software* Matlab, a partir da gravação dos dados em planilhas *excel* pelo *software* R.

4 APRESENTAÇÃO DOS RESULTADOS

Conforme foi apresentado no capítulo 3, o sistema de coleta de dados gerou uma base com mais de cem milhões de observações em um período definido de três anos, o qual necessitou de uma "limpeza" dos dados, realizada por uma rotina em *Visual Basic for Applications* (VBA). O processamento dessas rotinas em VBA para a limpeza dos relatórios de dados gerou uma elevada exigência computacional, sendo que em alguns comandos o tempo de execução ultrapassava uma hora de processamento de dados.

Após a limpeza dos dados, considerando a exclusão das observações dos equipamentos não selecionados e a eliminação de 46 variáveis não significativas para o OEE, restou uma magnitude de 712.536 observações, inseridas em 23 variáveis distintas, as quais são as candidatas a ser a base para a construção do modelo preditivo. Tendo em vista uma visualização de todo o processo realizado, o fluxo da FIGURA 33 visa resumir as etapas para a elaboração do modelo para posterior detalhamento nas seções subsequentes.

FIGURA 33 – SINTESE DO PROCESSO PARA CRIAÇAO DO MODELO PREDITIVO



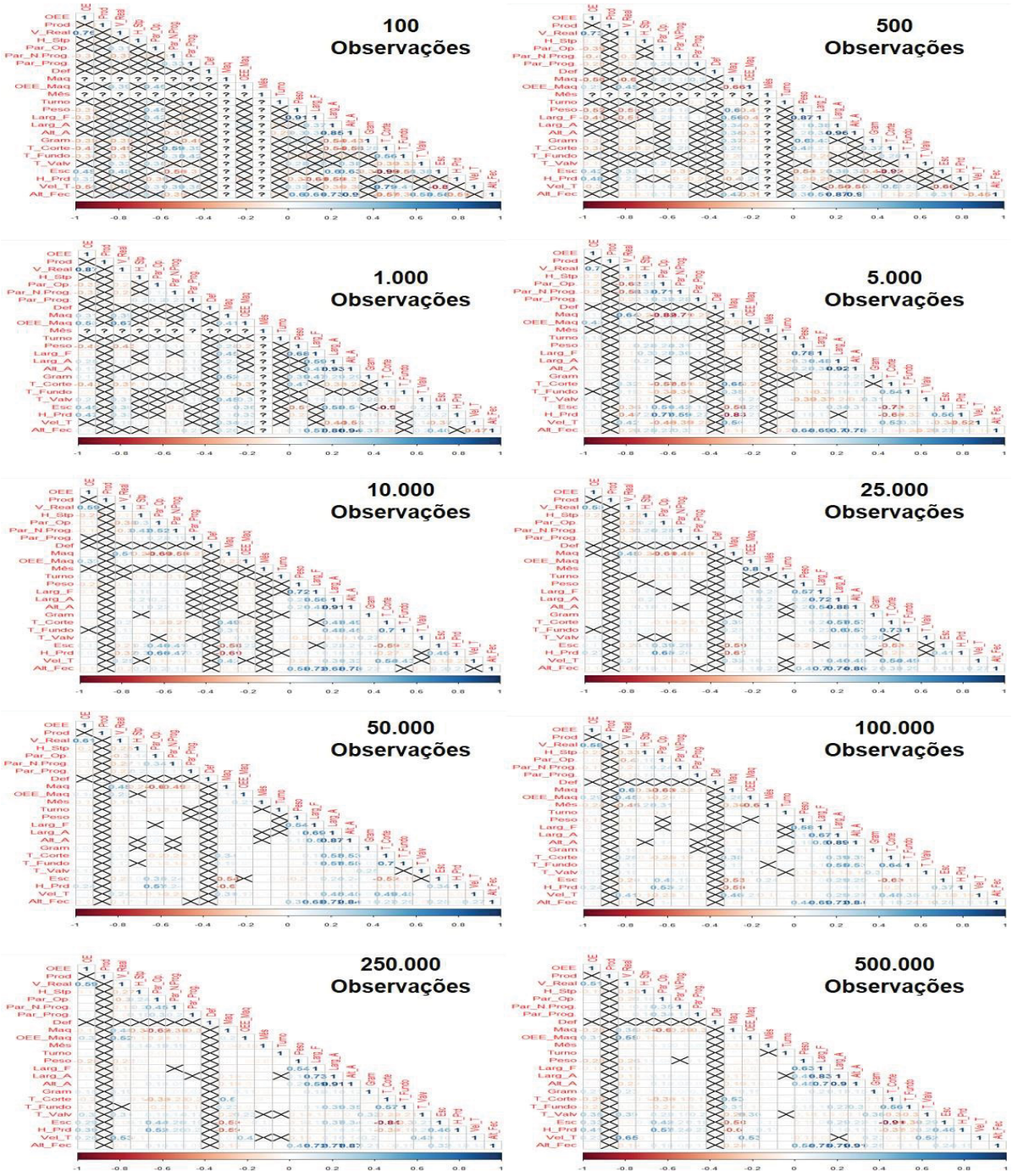
FONTE: O autor (2019).

4.1 ANÁLISE DAS MATRIZES DE CORRELAÇÕES

Tendo disponível o banco de dados com 23 variáveis e pouco mais de 712 mil observações, inseriu-se esses dados em um algoritmo no *software* R, o qual após as simulações realizadas, retornou os resultados dos p-valores para cada uma das situações. Na FIGURA 34 pode ser observado o comportamento da quantidade de p-valores significativos perante as variáveis presentes no banco de dados. Como resultado da análise o *software* R retorna uma matriz triangular comparando cada uma das variáveis. Nesta matriz existe para cada correlação um valor referente à

probabilidade estatística de correlação. Esses valores são classificados em uma escala de -1 (vermelhos) até +1 (azuis). Na matriz também estão presentes os p-valores, representados por um “X” nas correlações entre as variáveis. A presença do “X” indica p-valor acima de 5% (correlação não significativa estatisticamente), enquanto a ausência do “X” demonstra p-valor abaixo de 5% (correlação significativa estatisticamente).

FIGURA 34 – QUANTIDADE DE P-VALORES PELA QUANTIDADE DE OBSERVAÇÕES



FONTE: O autor (2019).

A TABELA 4 representa de forma resumida a quantidade de p-valores menores que 5%, e que apresentam uma correlação significativa em relação à quantidade de observações utilizadas na análise.

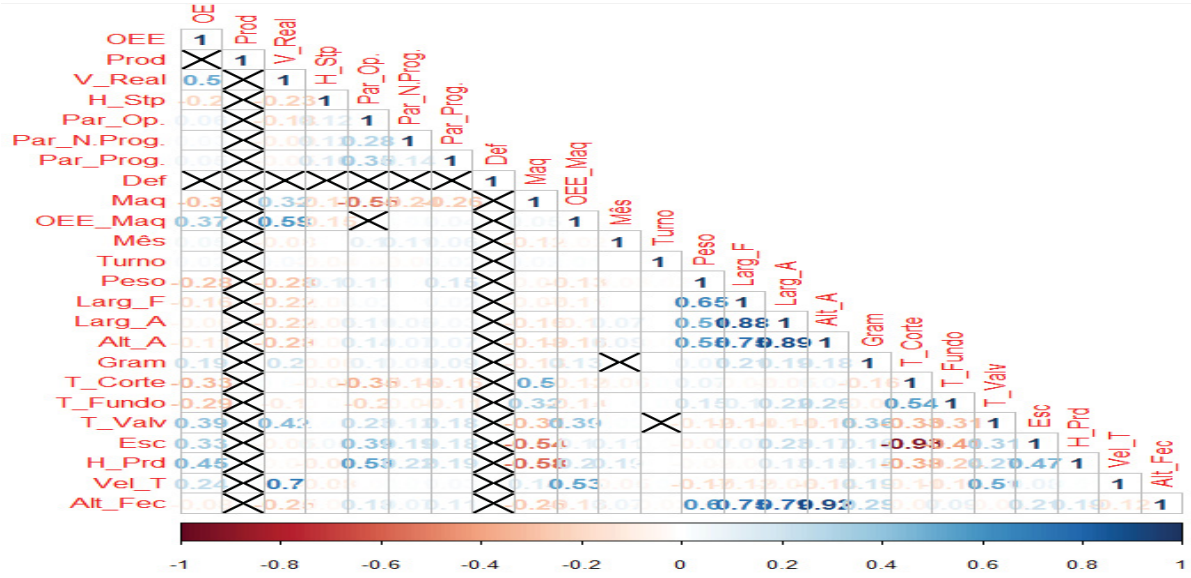
TABELA 4 – QUANTIDADE DE P-VALORES POR NÚMERO DE OBSERVAÇÕES

Número de Observações	Quantidade de p-valores < 5%
100	202
500	282
1.000	296
5.000	386
10.000	406
25.000	430
50.000	448
100.000	456
250.000	470
500.000	476
712.536	480

FONTE: O autor (2019).

Na matriz com 100 observações, presente na FIGURA 34, é verificada uma quantidade baixa de p-valores significativos, e conforme são aumentadas a quantidade de observações, os p-valores se tornam cada vez mais significativos e a quantidade de “X” na tabela é reduzida. Por fim, é observado na FIGURA 35 a quantidade máxima de observações presentes no banco de dados, sendo que as variáveis possuíam níveis de correlação significativa. As únicas variáveis que não demonstram correlação com as demais são as de Taxa de Produção e Defeitos. Tais variáveis podem possuir baixa correlação, pois em determinados momentos de produção são independentes das demais. Como as taxas de produção são elevadas, na ordem de milhares de unidades, quando existem eventos gerando defeitos ou baixa produção, os mesmo podem ser insignificantes perante o montante total produzido no mesmo período.

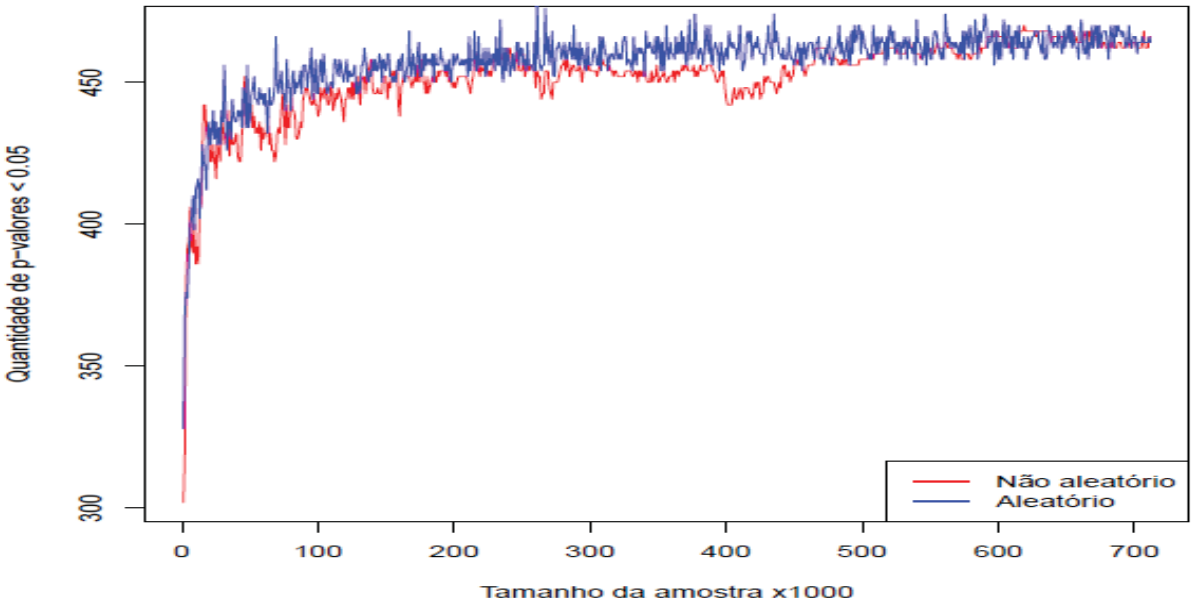
FIGURA 35 – QUANTIDADE DE P-VALORES EM 712.536 OBSERVAÇÕES



FONTE: O autor (2019).

Para observar de uma forma mais contínua o comportamento dos p-valores, realizou-se a mesma análise, porém com intervalos de 100 observações, conforme demonstrado na GRÁFICO 1.

GRÁFICO 1 – COMPORTAMENTO DO P-VALOR EM INTERVALOS DE 100 OBSERVAÇOES

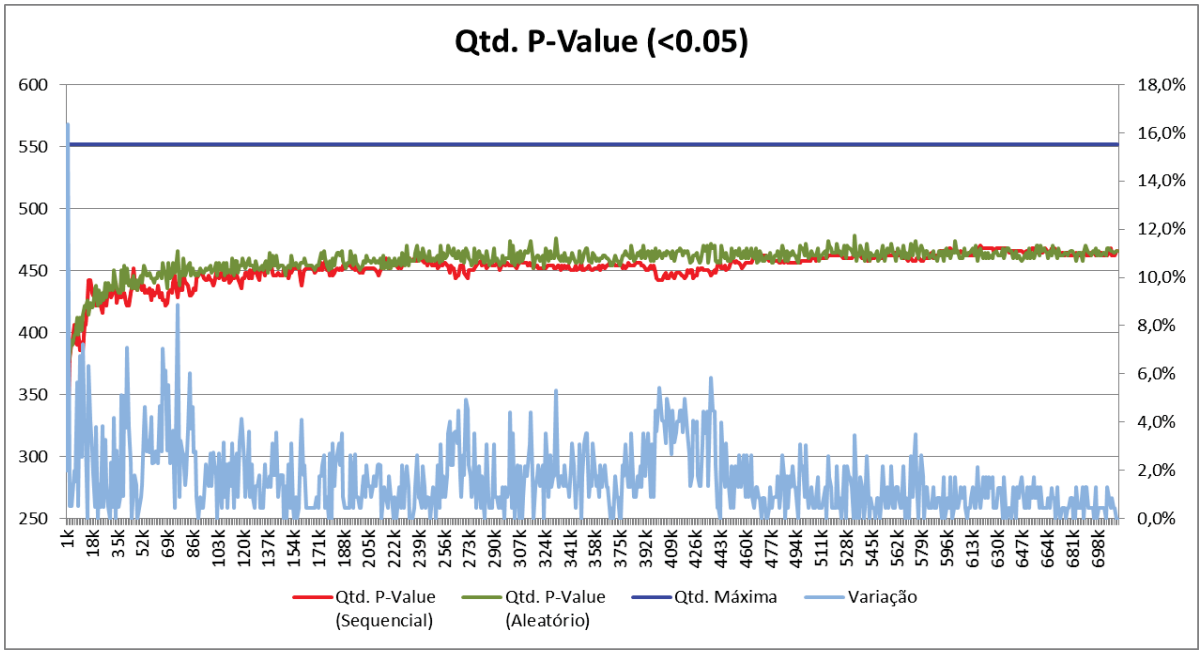


FONTE: O autor (2019).

No GRÁFICO 2 é possível verificar na linha inferior azul clara a taxa de variabilidade entre os *p-valores* aleatórios e não aleatórios, sendo que a linha reta superior em azul escuro, representa o limite máximo de p-valores significativos que

poderiam ser alcançados. Levando em consideração as 23 variáveis e mais o OEE poderia ser alcançado no máximo a quantidade de 576 p-valores significativos. Pelo gráfico é possível concluir que mesmo que se aumentasse a quantidade de observações, para essa amostra a quantidade de p-valores representativos continuaria próximo da ordem de 470.

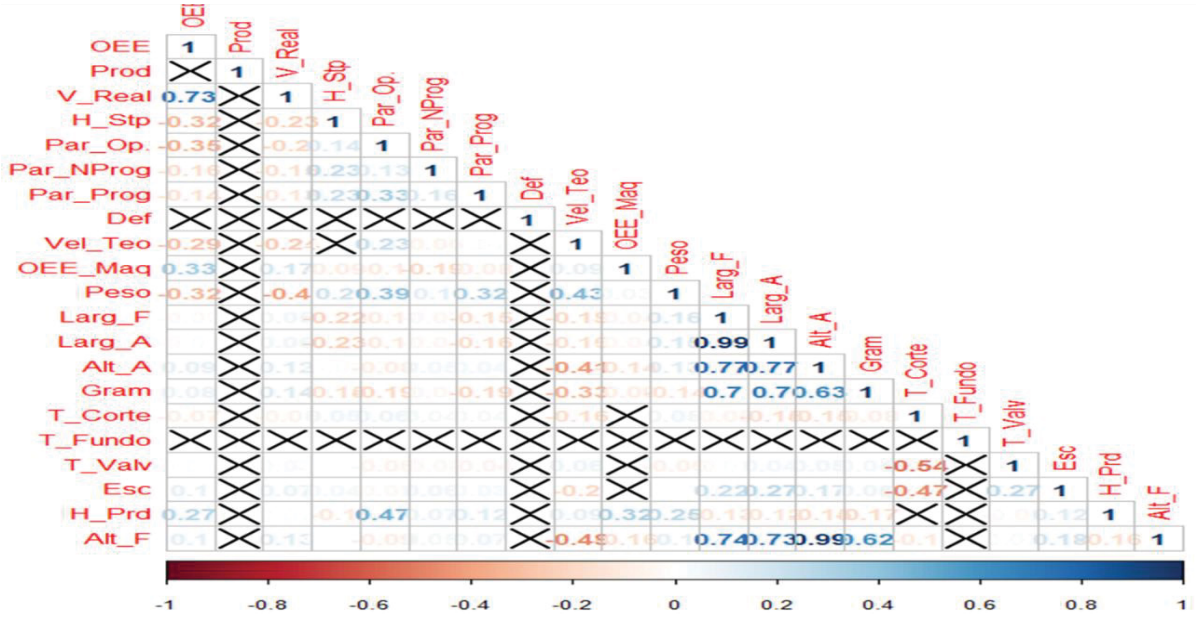
GRÁFICO 2 – VARIABILIDADE DO P-VALOR ENTRE OBSERVAÇÕES ALEATÓRIAS E NÃO-ALEATÓRIAS



FONTE: O autor (2019).

Para demonstrar o comportamento do p-valor sem a influência dos equipamentos, foi realizado as mesmas análises para os o três equipamentos selecionados M11, M12 e M14. Na FIGURA 36 é posivel observar o comportamento do p-valor para o equipamento M11, no qual existem cerca de 186.000 observações e uma quantidade de 440 p-valores significativos.

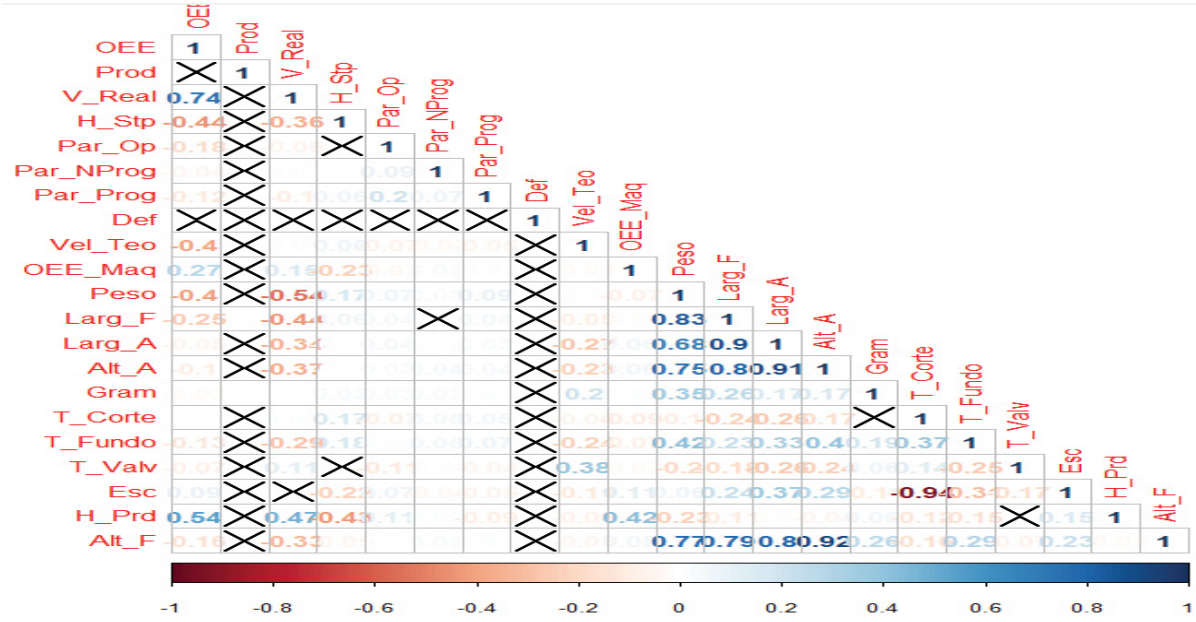
FIGURA 36 – COMPORTAMENTO DOS P-VALORES PARA O EQUIPAMENTO M11



FONTE: O autor (2019).

Na FIGURA 37 verifica-se para o equipamento M12, uma quantidade de 474 p-valores significativos para próximo de 186.000 observaçoes

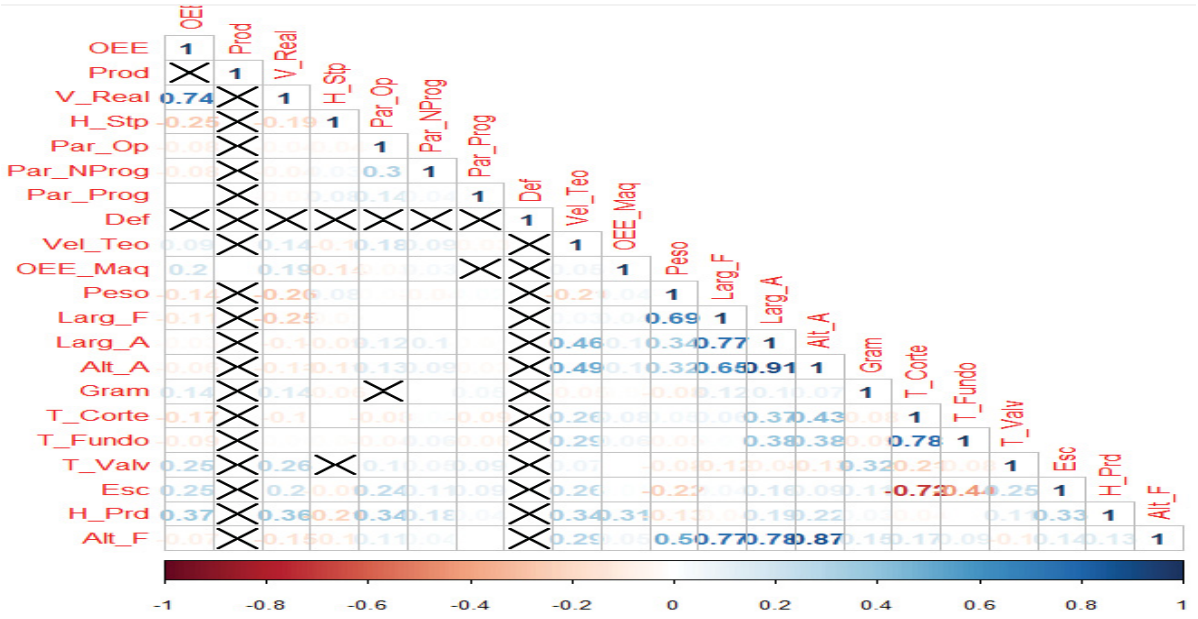
FIGURA 37 – COMPORTAMENTO DOS P-VALORES PARA O EQUIPAMENTO M12



FONTE: O autor (2019).

Para o equipamento M14, com 278.000 observações, verifica-se uma quantidade de 478 p-valores significativos, conforme FIGURA 38.

FIGURA 38 – COMPORTAMENTO DOS P-VALORES PARA O EQUIPAMENTO M14



FONTE: O autor (2019).

Analisando as tabelas por máquina é possível observar também que após 140 mil observações à quantidade de p-valores significativos são relevantes. Nestas imagens ainda é possível observar que as variáveis de Produção (Prod) e de Defeitos (Def) não obtiveram p-valores significativos para se afirmar que possuem correlação com as demais variáveis, como já demonstrado anteriormente. Desta forma, tais variáveis podem ser retiradas do modelo preditivo.

4.2 ESCOLHA DOS MÉTODOS ESTATÍSTICOS E MULTIVARIADOS – MODELO I

4.2.1 Análise de componentes principais – Modelo I

Como foram demonstradas fortes relações entre as variáveis por meio da análise de correlação, iniciou-se o processo para a análise de componentes principais. Porém duas variáveis demonstraram baixas correlações, as variáveis de produção (Prod) e Defeitos (Def) em relação às demais variáveis. Em virtude disso, optou-se por eliminar tais variáveis do processo de construção do modelo por estas serem pouco significativas.

Também foi demostrado que mais de 100 mil observações são pouco significativas para a criação do modelo, desta forma, visando ganho computacional,

optou-se por desenvolver os modelos com quantidades de observações próximas desse valor.

A TABELA 5 demonstra os autovalores das componentes principais para as 21 variáveis analisadas.

TABELA 5 – PESOS EM RELAÇÃO AOS TEMAS DE PESQUISA

Componentes Principais	Autovalores	Componentes Principais	Autovalores
CP1	4,5295	CP12	0,5295
CP2	3,8599	CP13	0,4719
CP3	2,4273	CP14	0,4548
CP4	1,2830	CP15	0,3491
CP5	1,1798	CP16	0,2735
CP6	1,0164	CP17	0,2627
CP7	0,9122	CP18	0,2181
CP8	0,8687	CP19	0,0883
CP9	0,8576	CP20	0,0237
CP10	0,7252	CP21	0,0146
CP11	0,6541	-	-

FONTE: O autor (2019).

A FIGURA 39 demonstra os autovetores (componentes principais), desvio padrão e variância gerados para cada variável analisada.

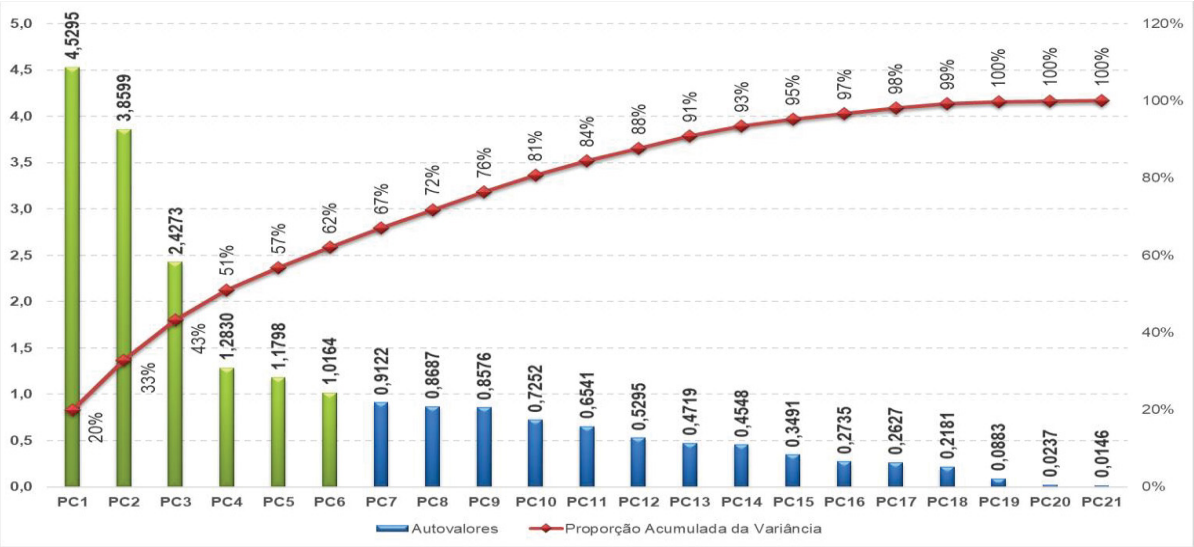
FIGURA 39 – AUTOVETORES DAS VARIÁVEIS

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21
V_Real	-0,24	0,18	-0,38	0,27	-0,06	-0,07	0,04	-0,05	0,07	-0,06	0,17	0,05	0,02	0,04	-0,11	-0,62	0,09	-0,49	0,03	0,00	-0,01
H_Stp	0,01	0,10	0,28	0,24	0,38	0,29	0,21	-0,06	0,03	-0,02	-0,34	0,51	0,32	-0,25	0,13	-0,10	0,02	-0,08	-0,02	0,02	0,00
Par_Op.	0,04	0,12	0,41	0,34	-0,33	-0,16	0,10	-0,04	0,13	-0,02	0,23	-0,15	0,11	0,15	0,64	0,01	0,13	-0,05	-0,01	-0,01	0,00
Par_N.Prog.	0,03	-0,01	0,12	0,26	0,39	-0,47	-0,70	0,12	-0,11	-0,10	-0,03	0,04	0,03	-0,04	0,02	-0,01	0,06	-0,01	0,00	0,01	0,00
Par_Prog.	0,01	0,14	0,09	0,15	0,08	0,29	0,14	0,51	-0,54	-0,46	0,14	-0,23	-0,09	-0,03	-0,02	0,00	0,03	-0,02	-0,01	0,00	0,00
Maq	-0,25	-0,20	-0,28	0,08	0,06	0,10	-0,08	-0,07	0,06	-0,17	0,26	-0,13	0,53	-0,15	0,14	0,29	-0,48	-0,09	-0,10	-0,05	0,03
OEE_Maq	-0,05	0,09	0,32	0,44	0,26	0,05	0,17	-0,17	0,29	0,07	0,12	-0,36	-0,17	0,06	-0,45	0,04	-0,26	0,08	0,08	-0,03	0,01
Mês	0,07	0,04	0,02	-0,12	0,32	-0,39	0,40	0,12	-0,33	0,47	0,41	0,08	0,22	0,03	-0,04	0,01	0,01	-0,01	0,01	-0,01	0,00
Turno	-0,02	-0,02	-0,04	-0,03	0,01	0,00	0,01	0,79	0,58	0,15	-0,01	0,04	0,05	-0,02	-0,01	-0,01	-0,01	0,01	0,00	0,00	0,00
Peso	0,37	-0,12	-0,05	0,07	-0,05	0,27	-0,19	0,04	-0,09	0,13	0,09	0,12	0,17	0,39	0,05	-0,28	-0,29	0,18	0,53	-0,15	0,02
Larg_F	0,35	-0,04	-0,12	0,16	0,02	0,31	-0,14	-0,08	0,12	0,04	0,38	0,12	0,07	0,03	-0,19	0,39	0,47	-0,25	0,01	0,22	-0,11
Larg_A	0,36	0,03	-0,23	0,04	0,05	-0,29	0,24	-0,04	0,13	-0,29	-0,07	-0,03	-0,08	-0,29	0,05	0,15	0,03	-0,09	0,38	-0,11	0,53
Alt_A	0,43	-0,04	-0,18	0,08	0,03	-0,18	0,16	-0,02	0,08	-0,17	-0,04	0,00	-0,10	-0,18	0,07	-0,05	-0,16	0,02	-0,13	-0,25	-0,73
Gram	0,11	0,22	-0,19	0,21	-0,14	0,14	-0,13	0,04	-0,17	0,49	-0,31	-0,47	0,22	-0,39	0,03	-0,02	0,13	0,05	0,06	-0,03	-0,02
T_Corte	-0,08	-0,48	-0,06	0,30	-0,12	-0,07	0,10	0,08	-0,14	0,13	-0,04	0,11	-0,27	-0,21	0,10	-0,01	-0,17	0,01	0,19	0,62	-0,09
T_Fundo	0,06	-0,37	-0,10	0,19	-0,01	-0,19	0,23	0,05	-0,05	-0,13	-0,43	-0,19	0,36	0,50	-0,17	0,05	0,24	-0,03	-0,13	0,03	0,01
T_Valv	-0,05	0,42	-0,21	0,19	-0,05	-0,05	0,01	0,09	-0,12	0,17	-0,25	0,22	-0,24	0,35	0,08	0,46	-0,31	-0,28	0,01	-0,02	-0,03
Esc	0,19	0,43	0,01	-0,29	0,10	-0,09	0,00	-0,06	0,12	-0,19	-0,07	-0,19	0,24	0,11	0,00	-0,11	-0,16	0,02	0,07	0,67	-0,11
H_Prd	0,07	0,12	0,26	0,08	-0,60	-0,22	-0,05	0,09	-0,11	-0,08	0,03	0,29	0,28	-0,19	-0,49	0,04	-0,14	0,01	0,00	-0,01	0,00
Vel_T	-0,19	0,26	-0,38	0,29	-0,02	-0,05	0,08	-0,04	0,05	-0,11	0,12	0,21	0,03	0,06	0,01	0,02	0,19	0,74	-0,03	0,06	0,00
Alt_Fec	0,44	-0,03	-0,08	0,13	-0,03	0,08	-0,05	0,01	-0,02	0,10	0,07	0,05	-0,11	0,03	0,03	-0,20	-0,24	0,06	-0,69	0,10	0,40
Desv. Padrão	2,06	1,66	1,49	1,28	1,11	1,05	1,03	1,00	0,99	0,96	0,88	0,84	0,82	0,73	0,63	0,56	0,53	0,50	0,34	0,21	0,08
Variância	4,22	2,74	2,22	1,65	1,23	1,10	1,07	1,01	0,98	0,92	0,78	0,71	0,68	0,53	0,40	0,31	0,28	0,25	0,11	0,04	0,01

FONTE: O autor (2019).

O GRÁFICO 3 demonstra em forma de barras a distribuição dos componentes principais, sendo que os em verde representam os elementos com autovalores acima de 0,7, os quais são utilizados para escolha das variáveis pelo método de Jolliffe B2 e B4. A linha em vermelho representa a variância acumulada.

GRÁFICO 3 – COMPONENTES PRINCIPAIS GERADOS



FONTE: O autor (2019).

Nas FIGURAS 40 e 41 são demonstradas as matrizes de correlação entre as variáveis originais e as componentes principais. Utilizou-se tais valores para a seleção das variáveis pelas técnicas de Jolliffe B2 e Jolliffe B4, respectivamente.

FIGURA 40 – SELEÇÃO DAS VARIÁVEIS PELO MÉTODO JOLLIFFE B2

Variáveis	JOLIFFE B2										
	CP11	CP12	CP13	CP14	CP15	CP16	CP17	CP18	CP19	CP20	CP21
V_Real	0,0420	0,0410	0,0795	0,1047	0,1830	0,0691	0,0898	0,3352	0,0195	0,0022	0,0002
H_Stp	0,0117	0,1377	0,0292	0,1791	0,0810	0,0381	0,0470	0,0019	0,0068	0,0011	0,0003
Par_Op.	0,0247	0,2090	0,4400	0,0730	0,1970	0,0291	0,1162	0,0256	0,0053	0,0006	0,0001
Par_N.Prog.	0,1326	0,0360	0,0463	0,0275	0,0180	0,0093	0,0057	0,0125	0,0011	0,0003	0,0002
Par_Prog.	0,3177	0,0777	0,1493	0,0638	0,0040	0,0325	0,0188	0,0126	0,0038	0,0013	0,0007
Maq	0,0814	0,0545	0,2826	0,0341	0,1360	0,0714	0,2277	0,2283	0,0214	0,0016	0,0022
OEE_Maq	0,1065	0,4463	0,0586	0,0317	0,2390	0,1077	0,0587	0,0428	0,0042	0,0003	0,0011
Mês	0,0929	0,0159	0,0595	0,0345	0,0300	0,0170	0,0048	0,0204	0,0011	0,0010	0,0006
Turno	0,0386	0,0066	0,0202	0,0040	0,0030	0,0153	0,0022	0,0027	0,0004	0,0000	0,0002
Peso	0,3214	0,0703	0,1776	0,3602	0,1790	0,0497	0,0964	0,0293	0,0765	0,0063	0,0047
Larg_F	0,1314	0,0199	0,0211	0,0305	0,1620	0,2700	0,0529	0,0420	0,1261	0,0038	0,0516
Larg_A	0,1183	0,0551	0,0475	0,1080	0,1110	0,1869	0,0010	0,0261	0,1279	0,0248	0,0680
Alt_A	0,1101	0,0638	0,0863	0,1195	0,0640	0,1595	0,0864	0,0275	0,0834	0,0932	0,0409
Gram	0,0487	0,2420	0,1722	0,0475	0,0270	0,0279	0,0698	0,0222	0,0267	0,0008	0,0001
T_Corte	0,1552	0,0689	0,1130	0,1589	0,0220	0,0007	0,0262	0,0232	0,0909	0,0750	0,0459
T_Fundo	0,4755	0,0466	0,0496	0,3071	0,0020	0,0299	0,0439	0,0468	0,0609	0,0007	0,0000
T_Valv	0,1066	0,2553	0,2348	0,2837	0,0200	0,1042	0,2191	0,0947	0,0026	0,0076	0,0013
Esc	0,2895	0,0385	0,0604	0,1091	0,0500	0,0315	0,0450	0,0174	0,0739	0,0735	0,0523
H_Prd	0,0775	0,1602	0,0148	0,1057	0,3210	0,1774	0,0500	0,1109	0,0316	0,0027	0,0007
Vel_T	0,0587	0,2881	0,0717	0,0915	0,0230	0,0839	0,2993	0,1454	0,0247	0,0050	0,0039
Alt_Fec	0,0338	0,0455	0,0930	0,0884	0,1060	0,2456	0,1186	0,0194	0,1502	0,0567	0,0279

FONTE: O autor (2019).

FIGURA 41 – SELEÇÃO DAS VARIÁVEIS PELO MÉTODO JOLLIFFE B4

Variáveis	JOLIFFE B4									
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
V_Real	0,4582	0,2949	0,7151	0,0375	0,0248	0,0207	0,0269	0,0869	0,0074	0,0323
H_Stp	0,0782	0,0473	0,3424	0,3853	0,5757	0,0972	0,3828	0,0587	0,0485	0,4129
Par_Op.	0,4065	0,5142	0,2573	0,3202	0,1217	0,0574	0,2701	0,0618	0,0097	0,0982
Par_N.Prog.	0,2002	0,2772	0,1574	0,3958	0,1516	0,2333	0,1738	0,0657	0,6579	0,3667
Par_Prog.	0,2352	0,3005	0,0872	0,5451	0,1717	0,2174	0,0977	0,1232	0,3747	0,4158
Maq	0,4929	0,5756	0,3461	0,0514	0,0410	0,0329	0,1223	0,1185	0,0017	0,2302
OEE_Maq	0,2189	0,4073	0,5856	0,0849	0,1296	0,0105	0,0531	0,3266	0,0115	0,1431
Mês	0,1505	0,1088	0,1251	0,1112	0,5383	0,2306	0,6260	0,0615	0,4265	0,0405
Turno	0,0306	0,0125	0,0247	0,0139	0,1625	0,9101	0,3203	0,0237	0,1901	0,0436
Peso	0,6277	0,3322	0,1412	0,1872	0,2448	0,0784	0,0578	0,1651	0,1475	0,0474
Larg_F	0,7777	0,3238	0,3203	0,1184	0,1329	0,0151	0,0647	0,0690	0,0208	0,0706
Larg_A	0,8404	0,2377	0,3171	0,1058	0,0508	0,0325	0,0752	0,0921	0,0701	0,0478
Alt_A	0,8640	0,2855	0,2453	0,0091	0,0967	0,0323	0,0083	0,0405	0,0516	0,0628
Gram	0,2425	0,2512	0,4289	0,0640	0,1217	0,0134	0,0044	0,7529	0,0279	0,0834
T_Corte	0,2822	0,7461	0,1445	0,4196	0,2030	0,0046	0,1456	0,1092	0,0314	0,0822
T_Fundo	0,0216	0,6138	0,1195	0,3240	0,2664	0,0609	0,0242	0,0294	0,1533	0,2648
T_Valv	0,0421	0,6719	0,3843	0,1984	0,1325	0,0060	0,0339	0,2086	0,0371	0,1253
Esc	0,4202	0,7044	0,1132	0,3861	0,0680	0,0311	0,1421	0,1169	0,0715	0,0232
H_Prd	0,3620	0,5932	0,0933	0,0541	0,4470	0,0441	0,2607	0,0200	0,0218	0,2043
Vel_T	0,2584	0,3903	0,6957	0,1305	0,0857	0,0593	0,1245	0,1694	0,0391	0,0662
Alt_Fec	0,8785	0,1687	0,2557	0,0245	0,0002	0,0039	0,0149	0,0367	0,0066	0,0051

FONTE: O autor (2019).

As variáveis selecionadas pelos métodos de Jolliffe B2 e B4, estão representadas no QUADRO 3.

QUADRO 3 – PARÂMETROS DE BUSCA

Método Jolliffe B2	Método Jolliffe B4	Método Jolliffe B2 e B4
V_Real	H_Stp	H_Stp
H_Stp	Par_N.Prog	Par_N.Prog
Par_N.Prog	Par_Prog	Par_Prog
Par_Prog	Maq	Mês
Mês	Mês	Turno
Turno	Turno	Gram
Gram	Gram	T_Corte
T_Corte	T_Corte	-
T_Fundo	T_Valv	-
Alt_Fec	Esc	-

FONTE: O autor (2019).

Após a escolha das variáveis realizou-se a análise de regressão linear múltipla nos equipamentos selecionados. O processo de seleção dos equipamentos levou em consideração as observações realizadas em relação aos períodos de fabricação, mais especificamente nos meses e turnos de trabalho, visando entender as interferências de mão-de-obra e sazonalidade, porém verificou-se poucas variações de observações

(apontamentos) em relação a esses quesitos, conforme demonstrado nas TABELAS 6 e 7.

TABELA 6 – PERCENTUAL DE OBSERVAÇÕES POR PERÍODO

Mês	Observações (%)	Mês	Observações (%)
1	8,76	7	7,92
2	5,52	8	8,59
3	9,92	9	8,95
4	8,82	10	9,27
5	8,46	11	9,22
6	7,34	12	7,23

FONTE: O autor (2019).

TABELA 7 – PERCENTUAL DE OBSERVAÇÕES POR TURNO DE TRABALHO

Turno	Observações (%)
1	34,18
2	33,23
3	32,59

FONTE: O autor (2019).

Porém, ao verificar as variações de observações por máquina, conforme demonstrado na TABELA 8, entendeu-se que alguns equipamentos possuíam um maior mix de produção, os quais poderiam contribuir para um modelo de regressão mais assertivo por serem mais generalistas. Assim, optou-se por desenvolver as equações de regressão nos equipamentos M11, M12 e M14.

TABELA 8 – PERCENTUAL DE OBSERVAÇÕES POR MÁQUINA

Máquina	Observações (%)
M8	0,30
M9	6,01
M10	7,81
M11	19,61
M12	26,25
M13	0,98
M14	39,04

FONTE: O autor (2019).

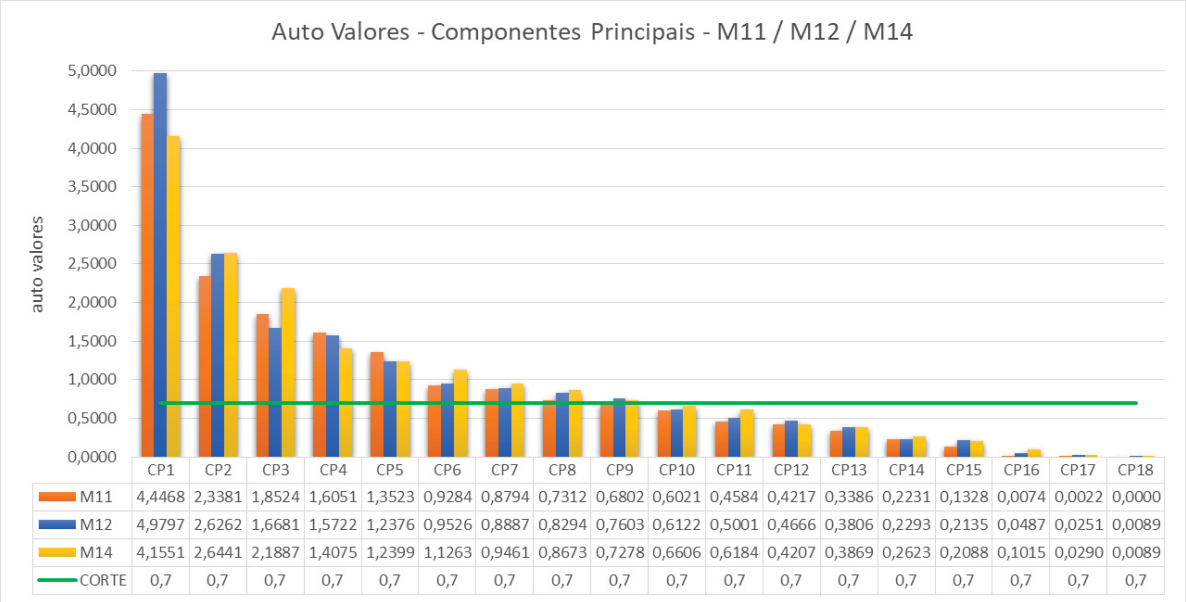
4.3 ESCOLHA DOS MÉTODOS ESTATÍSTICOS E MULTIVARIADOS – MODELO II

Nesta fase, foram desconsideradas as variáveis de Produção (Prod) e Defeitos (Def), pois as mesmas apresentaram baixa correlação. Também foram retiradas as variáveis mês e turno, por não apresentarem diferenças significativas em relação as observações e por serem um desdobramento das demais variáveis. A variável de máquina está sendo considerada, pois neste caso ela é constate em virtude da análise estar sendo realizada por esse parâmetro, restando assim 18 variáveis objetos de estudo para o modelo II.

4.3.1 Análise de componentes principais – Modelo II

Foi realizada a análise de componentes principais para cada uma das máquinas selecionadas de forma independente, sendo os resultados apresentados pela imagem do GRÁFICO 4.

GRÁFICO 4 – COMPONENTES PRINCIPAIS – M11 / M12 / M14



	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10	CP11	CP12	CP13	CP14	CP15	CP16	CP17	CP18
M11	4,4468	2,3381	1,8524	1,6051	1,3523	0,9284	0,8794	0,7312	0,6802	0,6021	0,4584	0,4217	0,3386	0,2231	0,1328	0,0074	0,0022	0,0000
M12	4,9797	2,6262	1,6681	1,5722	1,2376	0,9526	0,8887	0,8294	0,7603	0,6122	0,5001	0,4666	0,3806	0,2293	0,2135	0,0487	0,0251	0,0089
M14	4,1551	2,6441	2,1887	1,4075	1,2399	1,1263	0,9461	0,8673	0,7278	0,6606	0,6184	0,4207	0,3869	0,2623	0,2088	0,1015	0,0290	0,0089
CORTE	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7

FONTE: O autor (2019).

Como resultado da análise, observa-se que todas as máquinas possuem oito componentes principais com autovalores maiores do que 0,7, sendo que os equipamentos M12 e M14 apresentam uma nona componente com autovalor maior do que 0,7.

Analogamente a análise efetuada no modelo I, foi realizada a seleção das variáveis mais significativas em relação aos componentes principais, sendo utilizados novamente os métodos de Jolliffe B2 e B4. No QUADRO 4 é demonstrada uma síntese do resultado desse método para cada um dos equipamentos.

QUADRO 4 – JOLLIFFE B2 & B4 – M11 / M12 / M14

Máquina 11 (M11)		Máquina 12 (M12)		Máquina 14 (M14)	
Jolliffe B2	Jolliffe B4	Jolliffe B2	Jolliffe B4	Jolliffe B2	Jolliffe B4
H_Stp	H_Stp	V_Real	Par_Op	H_Stp	H_Stp
Par_Prog	Par_NProg	Par_NProg	Par_NProg	Par_Op	Par_Op
OEE_Maq	Vel_Teo	Par_Prog	Par_Prog	Par_Prog	Par_Prog
Larg_F	OEE_Maq	Vel_Teo	Vel_Teo	OEE_Maq	OEE_Maq
T_Valv	Peso	OEE_Maq	OEE_Maq	Gram	Alt_A
Esc	Alt_A	Alt_A	Alt_A	T_Corte	Gram
H_Prod	T_Corte	Gram	Esc	T_Fundo	T_Fundo
Alt_F	Esc	T_Corte	H_Prd	Esc	Esc

FONTE: O autor (2019).

Após a aplicação da técnica de Jolliffe, foi aplicada a análise de regressão linear múltipla, sendo esta realizada em cada um dos equipamentos selecionados, conforme demonstrado na FIGURA 42.

FIGURA 42 – RESULTADOS ANÁLISE REGRESSÃO MÚLTIPLA – M11 / M12 / M14

M11 – Variáveis de Jolliffe B2

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.59259 -0.06843  0.01579  0.07544  0.65498

Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
Alt_F     0.00020108  0.00002018   9.963   <2e-16 ***
Esc       0.00107324  0.00046015   2.332   0.0197 *
H_Prd     0.03035753  0.00222421  13.649   <2e-16 ***
H_Stp    -0.03563653  0.00106283 -33.530   <2e-16 ***
Larg_F   -0.00048257  0.00003643 -13.245   <2e-16 ***
OEE_Maq   0.00919320  0.00074506  12.339   <2e-16 ***
Par_Prog -0.00097566  0.00008207 -11.888   <2e-16 ***
T_Valv    0.00096060  0.00002452  39.181   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1091 on 19992 degrees of freedom
Multiple R-squared:  0.9563,    Adjusted R-squared:  0.9563
F-statistic: 5.466e+04 on 8 and 19992 DF,  p-value: < 2.2e-16
```

M11 – Variáveis de Jolliffe B4

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.65044 -0.05546  0.00554  0.07457  0.62307

Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
Alt_A     0.00028706  0.00001579  18.184  < 2e-16 ***
Esc       0.00855282  0.00037636  22.725  < 2e-16 ***
H_Stp    -0.02982854  0.00102965 -28.970  < 2e-16 ***
OEE_Maq   0.00718334  0.00079436   9.043  < 2e-16 ***
Par_NProg -0.01923497  0.00126763 -15.174  < 2e-16 ***
Peso      -0.00116386  0.00003734 -31.168  < 2e-16 ***
T_Corte    0.01900090  0.00045432  41.823  < 2e-16 ***
Vel_Teo   -0.00032901  0.00004135  -7.956 1.87e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1063 on 19992 degrees of freedom
Multiple R-squared:  0.9585,    Adjusted R-squared:  0.9585
F-statistic: 5.775e+04 on 8 and 19992 DF,  p-value: < 2.2e-16
```

M12 – Variáveis de Jolliffe B2

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.69055 -0.04373  0.00601  0.05208  0.30690

Coefficients:
              Estimate   Std. Error t value Pr(>|t|)
V_Real     0.005310814  0.000029142 182.240  < 2e-16 ***
Par_NProg -0.000586819  0.000081949  -7.161 8.30e-13 ***
Par_Prog  -0.001301296  0.000082686 -15.738  < 2e-16 ***
Vel_Teo   -0.002291449  0.000024424 -93.818  < 2e-16 ***
OEE_Maq    0.001260157  0.000715528   1.761   0.0782 .
Alt_A      0.000022045  0.000004578   4.815 1.48e-06 ***
Gram       0.002889402  0.000087705  32.944  < 2e-16 ***
T_Corte    0.000427180  0.000069729   6.126 9.16e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09236 on 19992 degrees of freedom
Multiple R-squared:  0.9411,    Adjusted R-squared:  0.9411
F-statistic: 3.994e+04 on 8 and 19992 DF,  p-value: < 2.2e-16
```

M12 – Variáveis de Jolliffe B4

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.87925 -0.11341 -0.01542  0.12111  0.72558

Coefficients:
              Estimate   Std. Error t value     Pr(>|t|)
Par_Op    -0.287401522  0.011262604 -25.518      < 2e-16 ***
Par_NProg -0.001513749  0.000155823  -9.715      < 2e-16 ***
Par_Prog  -0.000122563  0.000157200  -0.780        0.436
Vel_Teo    0.000490142  0.000024625  19.904      < 2e-16 ***
OEE_Maq    0.029297481  0.001482860  19.757      < 2e-16 ***
Alt_A      0.000343428  0.000005362  64.044      < 2e-16 ***
Esc        0.000562058  0.000100205   5.609 0.0000000206 ***
H_Prd      0.147148720  0.006209125  23.699      < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1759 on 19991 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7865,    Adjusted R-squared:  0.7864
F-statistic:  9204 on 8 and 19991 DF,  p-value: < 2.2e-16
```

M14 – Variáveis de Jolliffe B2

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.64648 -0.09149  0.00696  0.08579  0.64726

Coefficients:
              Estimate  Std. Error t value Pr(>|t|)
H_Stp     -0.07864678  0.00252774 -31.113  < 2e-16 ***
Par_Op    -0.28358195  0.01046497 -27.098  < 2e-16 ***
Par_Prog  -0.00078509  0.00013470  -5.828 5.69e-09 ***
OEE_Maq    0.03263544  0.00173730  18.785  < 2e-16 ***
Gram       0.00345663  0.00009337  37.020  < 2e-16 ***
T_Corte    0.00007569  0.00019982   0.379    0.705
T_Fundo    0.00099823  0.00014221   7.020 2.30e-12 ***
Esc        0.00382887  0.00015348  24.946  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1268 on 19992 degrees of freedom
Multiple R-squared:  0.8808,    Adjusted R-squared:  0.8807
F-statistic: 1.846e+04 on 8 and 19992 DF,  p-value: < 2.2e-16
```

M14 – Variáveis de Jolliffe B4

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.65834 -0.08838  0.00800  0.08648  0.66365

Coefficients:
              Estimate   Std. Error t value Pr(>|t|)
H_Stp     -0.084513568  0.002525287  -33.47  < 2e-16 ***
Par_Op    -0.275402220  0.010391092  -26.50  < 2e-16 ***
Par_Prog  -0.000874792  0.000133750   -6.54 6.28e-11 ***
OEE_Maq    0.032312462  0.001724362   18.74  < 2e-16 ***
Alt_A     -0.000105327  0.000006129  -17.19  < 2e-16 ***
Gram       0.003452865  0.000090516   38.15  < 2e-16 ***
T_Fundo    0.002100404  0.000127123   16.52  < 2e-16 ***
Esc        0.004376914  0.000107272   40.80  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1259 on 19992 degrees of freedom
Multiple R-squared:  0.8825,    Adjusted R-squared:  0.8824
F-statistic: 1.877e+04 on 8 and 19992 DF,  p-value: < 2.2e-16
```

FONTE: O autor (2019).

No QUADRO 5 é possível verificar os coeficientes das equações referentes às variáveis de X_1 a X_8, geradas em cada uma das máquinas.

QUADRO 5 – COEFICIENTES EQUAÇÃO REGRESSÃO – M11 / M12 / M14

Variável	Máquina 11 (M11)		Máquina 12 (M12)		Máquina 14 (M14)	
Método	Jolliffe B2	Jolliffe B4	Jolliffe B2	Jolliffe B4	Jolliffe B2	Jolliffe B4
X_1	Alt_F	Alt_A	V_Real	Par_Op	H_Stp	H_Stp
X_2	Esc	Esc	Par_N.Prog	Par_N.Prog	Par_Op	Par_Op
X_3	H_Prd	H_Stp	Par_Prog	Par_Prog	Par_Prog	Par_Prog
X_4	H_Stp	OEE_Maq	Vel_Teo	Vel_Teo	OEE_Maq	OEE_Maq
X_5	Larg_F	Par_N.Prog	OEE_Maq	OEE_Maq	Gram	Alt_A
X_6	OEE_Maq	Peso	Alt_A	Alt_A	T_Corte	Gram
X_7	Par_Prog	T_Corte	Gram	Esc	T_Fundo	T_Fundo
X_8	T_Valv	Vel_Teo	T_Corte	H_Prd	Esc	Esc

FONTE: O autor (2019).

Para o equipamento M11 foi determinada a equação (1.36), sendo os coeficientes demonstrados pelo QUADRO 5, para as variáveis originárias de Jolliffe B2 e a equação (1.37) para as de Jolliffe B4, tais equações retornaram um R^2 ajustado de 0,9563 e 0,9585 respectivamente.

$$Y(X) = 0{,}0002X_1 + 0{,}0011X_2 + 0{,}0303X_3 - 0{,}0356X_4 \\ -0{,}0005X_5 + 0{,}0092X_6 - 0{,}0010X_7 + 0{,}0010X_8 + 0{,}1091 \quad (1.36)$$

$$Y(X) = 0{,}0003X_1 + 0{,}0086X_2 - 0{,}0298X_3 + 0{,}0072X_4 \\ -0{,}0192X_5 - 0{,}0012X_6 + 0{,}0190X_7 + 0{,}0003X_8 + 0{,}1063 \quad (1.37)$$

No equipamento M12 as equações (1.38) e (1.39) representam os modelos para Jolliffe B2 e B4, com um R^2 ajustado de 0,9411 e 0,7564 respectivamente.

$$Y(X) = 0{,}0053X_1 - 0{,}0006X_2 - 0{,}0013X_3 - 0{,}0023X_4 \\ +0{,}0013X_5 + 0{,}00002X_6 + 0{,}0029X_7 + 0{,}0004X_8 + 0{,}0924 \quad (1.38)$$

$$Y(X) = -0{,}2874X_1 - 0{,}0015X_2 - 0{,}0001X_3 + 0{,}0005X_4 \\ +0{,}0293X_5 + 0{,}0003X_6 + 0{,}0006X_7 + 0{,}1471X_8 + 0{,}1759 \quad (1.39)$$

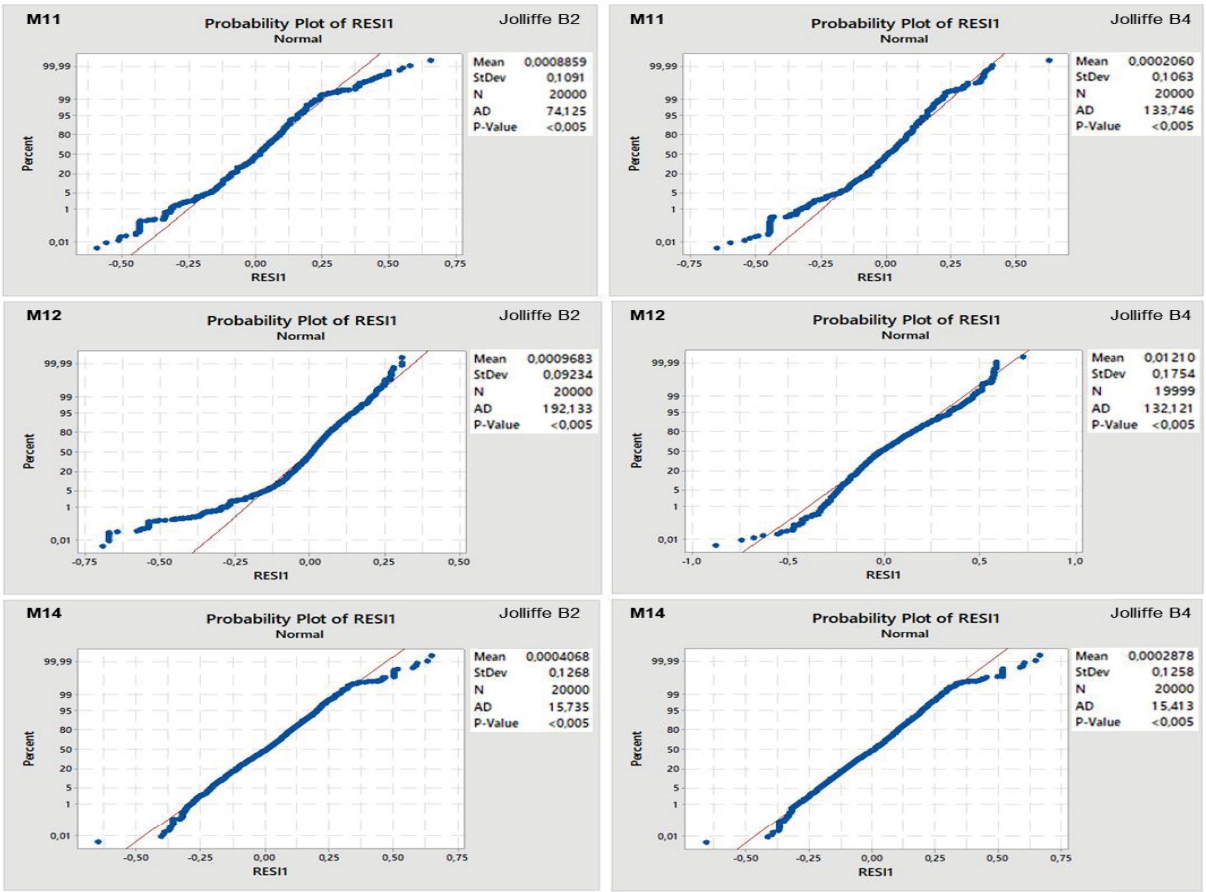
Com um R^2 ajustado de 0,8807 e 0,8824 para Jolliffe B2 e B4 respectivamente, as equações (1.40) e (1.41) demonstram as equações de regressão para o equipamento M14.

$$Y\,(X) = -0{,}0786X_1 - 0{,}2835X_2 - 0{,}0008X_3 + 0{,}0326X_4 \\ +0{,}0035X_5 + 0{,}00008X_6 + 0{,}0010X_7 + 0{,}0038X_8 + 0{,}1268 \quad (1.40)$$

$$Y\,(X) = -0{,}0845X_1 - 0{,}2754X_2 - 0{,}0009X_3 + 0{,}0323X_4 \\ -0{,}0001X_5 + 0{,}0035X_6 + 0{,}0021X_7 + 0{,}0044X_8 + 0{,}1259 \quad (1.41)$$

Todas as equações geradas pelo método de regressão linear múltipla apresentaram um R^2 elevado, o que pode-se dizer que as mesmas explicam os modelos de forma satisfatória. Porém, para afirmar que tais modelos são estatisticamente válidos, efetuou-se a análise dos resíduos, começando pelo teste de normalidade dos resíduos de cada um dos modelos. Os resultados de tais testes estão apresentados nos gráficos contidos na FIGURA 43.

FIGURA 43 – RESULTADOS DOS TESTES DE NORMALIDADE – M11 / M12 / M14



FONTE: O autor (2019).

O teste de normalidade usado foi o de Anderson-Darling. Como o p-valor foi menor que o nível de significância de 5%, rejeita-se a hipótese de normalidade dos resíduos, desta forma não atendendo ao pressuposto de normalidade dos mesmos. Pelo fato de todos os modelos não atenderem ao pressuposto de normalidade dos resíduos, não se efetuaram os testes de homocedasticidade e independência dos mesmos. O não atendimento do pressuposto de normalidade inviabiliza a realização de inferências dos coeficientes da equação de regressão em relação ao OEE e, em virtude disso, outra técnica para a obtenção do modelo preditivo foi desenvolvida.

4.4 ESCOLHA DOS MÉTODOS ESTATÍSTICOS E MULTIVARIADOS – MODELO III

Pelo fato de não se poder realizar inferências nos modelos de regressão estimados optou-se por uma outra técnica preditiva, as redes neurais artificiais (RNA).

4.4.1 Redes Neurais Artificiais - RNA

Para a família de produtos definida, foram treinadas as redes neurais para dois produtos distintos (Produto A e Produto B), com o intuito de comparar dois elementos diferentes, onde foram calculados os erros de previsão para os conjuntos de testes, conforme demonstrado nas TABELAS 9 e 10.

TABELA 9 – ERROS DE PREVISÃO – REDES PRODUTO A

Número de neurônios na camada oculta	Erro médio das simulações (15 eventos)	Número de neurônios na camada oculta	Erro médio das simulações (15 eventos)
1	0,06344	11	0,00717
2	0,06549	12	0,01345
3	0,04695	13	0,00846
4	0,04840	14	0,01331
5	0,06199	15	0,01281
6	0,03207	16	0,00672
7	0,02349	17	**0,00518**
8	0,02126	18	0,01013
9	0,01334	19	0,00613
10	0,01128	20	0,00629

FONTE: O autor (2019).

Observando a TABELA 9, verifica-se que a rede que possui menor erro médio é a rede com 17 neurônios na camada oculta para o produto A.

TABELA 10 – ERROS DE PREVISÃO – REDES PRODUTO B

Número de neurônios na camada oculta	Erro médio das simulações (15 eventos)	Número de neurônios na camada oculta	Erro médio das simulações (15 eventos)
1	0,04764	11	0,04189
2	0,04480	12	**0,00291**
3	0,05287	13	0,01019
4	0,03626	14	0,02976
5	0,05014	15	0,01085
6	0,04677	16	0,04494
7	0,03830	17	0,00709
8	0,03697	18	0,00575
9	0,03555	19	0,01342
10	0,01661	20	0,00980

FONTE: O autor (2019).

De forma análoga ao produto A, observando a TABELA 10, referente ao produto B, a rede que possui menor erro médio é a rede com 12 neurônios na camada oculta.

Concluído as análises da RNA, seguiu-se para a próxima etapa de validação das redes, onde após a finalização dos cálculos, foram calculados os erros médios para os 5 conjuntos de testes considerando a técnica *k-fold* para o produto A e para o produto B, conforme demonstrado na TABELA 11.

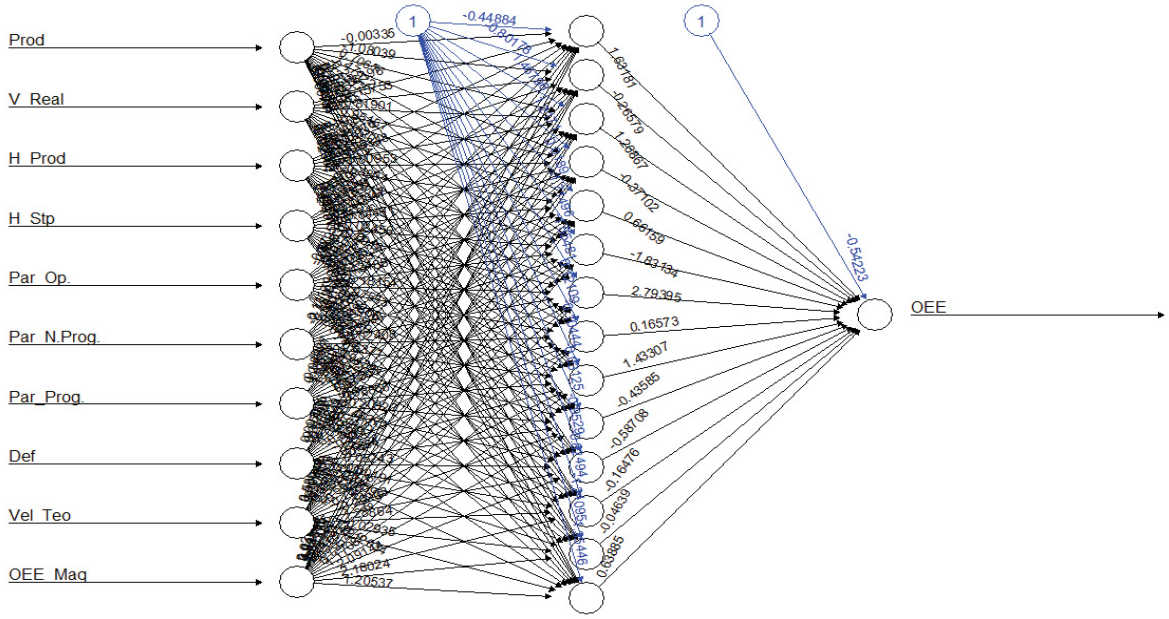
TABELA 11 – ERROS MÉDIOS – PRODUTO A E PRODUTO B

Erros gerados – Produto A	Erros gerados – Produto B
0,000001	0.020446
0,015918	**0.030000**
0,000011	0.008124
0,002161	0.004361
0,000004	0.029100

FONTE: O autor (2019).

Considerando um erro de 3%, foram julgadas como válidas as redes para os produtos A e B, pois nenhuma das simulações erraram mais do que essa taxa. Tendo a validação das duas redes, geraram-se os pesos finais de cada uma delas, treinando-as agora com todos os respectivos dados do conjunto (treinamento e teste juntos). Concluindo essas etapas, tem-se as redes finalizadas, conforme exemplo demonstrado na FIGURA 44.

FIGURA 44 – EXEMPLO DE REDE NEURAL GERADA

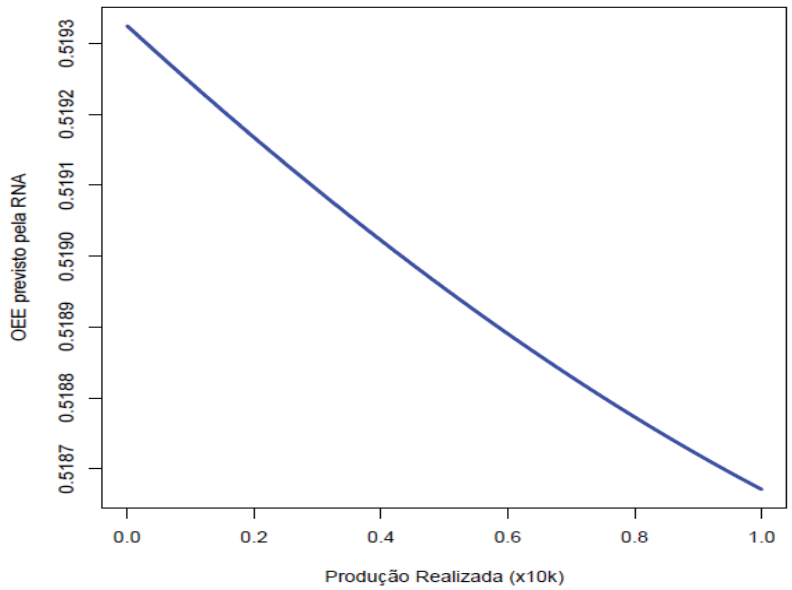


FONTE: O autor (2019).

Para verificar os resultados de saída gerados pela rede e analisar o comportamento do OEE perante a influência das variáveis, foram criados gráficos para análise de previsão para o produto A, devido ao menor erro apresentado nos resultados de saída da rede. Nos gráficos de 5 a 13, podem ser observados os gráficos em 2D de predição do OEE para cada uma das variáveis de entrada para o produto A. A interpretação de tais gráficos permite verificar o comportamento do OEE conforme são variadas as entradas da rede neural artificial.

No GRÁFICO 5 pode ser observada a variação da taxa de produção realizada pelo OEE previsto. Apesar de aparentar uma queda no valor do OEE, a escala do gráfico está reduzida, sendo que a variação entre o maior valor de OEE e o menor é de apenas 0,0006, ou 0,06% de OEE. Nesta situação, percebe-se praticamente uma constante para os valores de OEE.

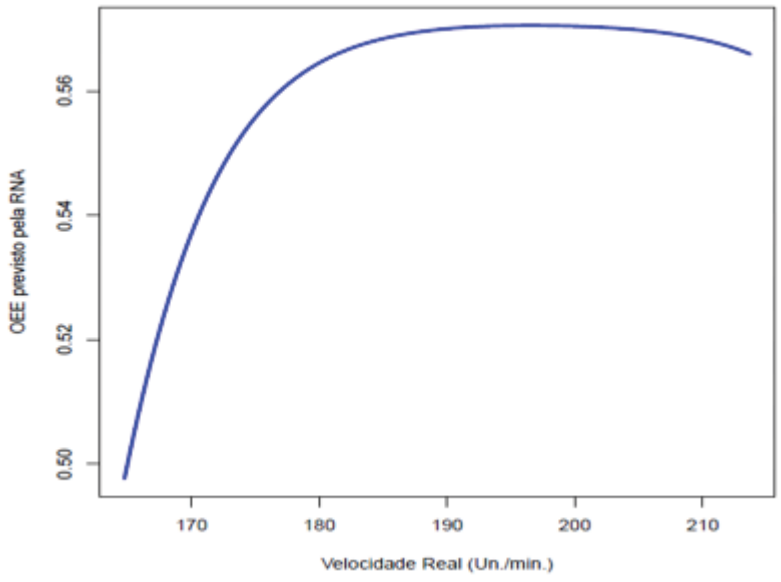
GRÁFICO 5 – OEE PREVISTO X TAXA DE PRODUÇÃO REALIZADA



FONTE: O autor (2019).

Analisando o GRÁFICO 6, observa-se um aumento das taxas de OEE, conforme aumentam-se as taxas de velocidade do equipamento. Essa conclusão não é uma surpresa considerando-se o OEE e os aumentos nas taxas de performance, porém é interessante observar que a partir de uma certa taxa de velocidade o OEE se estabiliza, não sendo mais significativo o aumento de velocidade. Para esse exemplo fica claro que aumentar a velocidade em muitos casos não necessariamente corresponde a um aumento de eficiência de um equipamento.

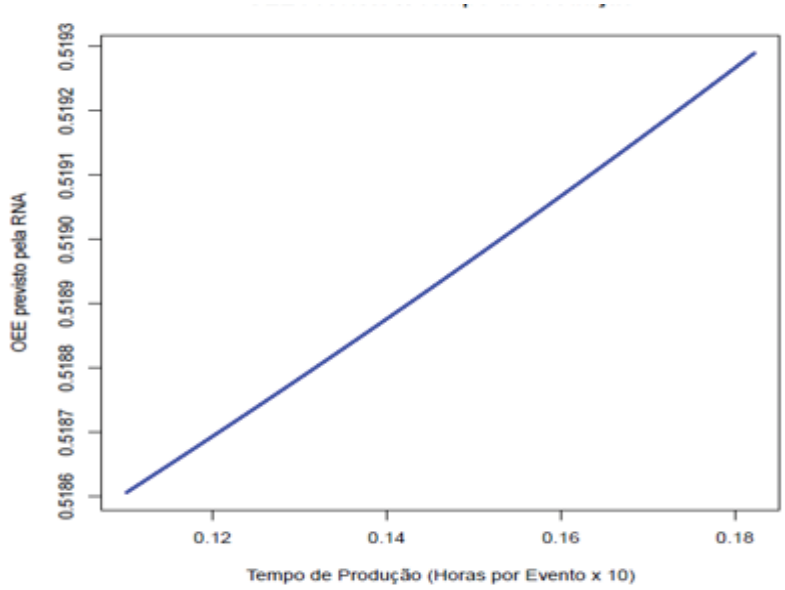
GRÁFICO 6 – OEE PREVISTO X VARIAÇÃO DA VELOCIDADE REAL



FONTE: O autor (2019).

Para o GRÁFICO 7 existe a mesma interpretação do OEE previsto para as taxas de produção. Neste cenário o OEE é praticamente constante, variando na ordem de 0,07%. Isso reforça que o tempo de produção de um produto não é um fator primordial para determinar a eficiência de um equipamento.

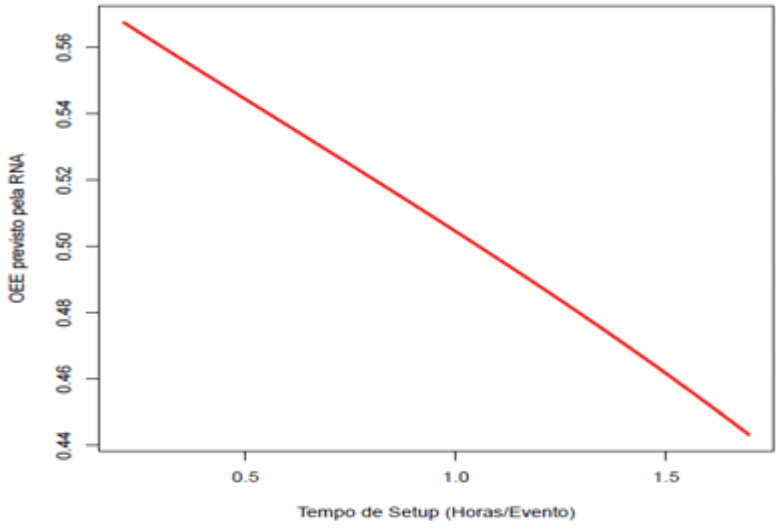
GRÁFICO 7 – OEE PREVISTO X TEMPO DE PRODUÇÃO



FONTE: O autor (2019).

No GRÁFICO 8, o tempo de *setup* pela previsão do OEE é demonstrado como o aumento do tempo de trocas, ou do número de trocas pode impactar na eficiência do equipamento. Neste caso, o OEE está sendo impactado diretamente pela perda de disponibilidade da máquina pelo *setup*.

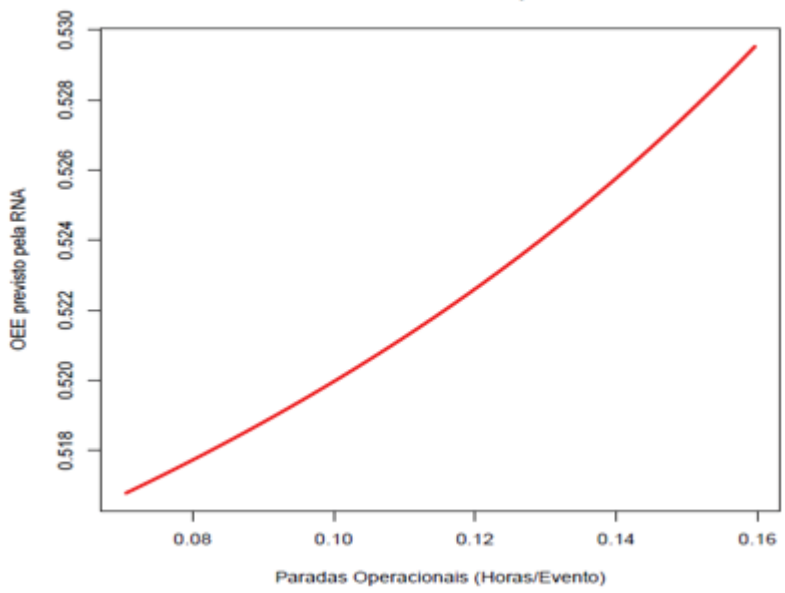
GRÁFICO 8 – OEE X TEMPO DE *SETUP*



FONTE: O autor (2019).

No GRÁFICO 9, observa-se que o aumento das paradas operacionais é seguido por um aumento de OEE. Se a análise for realizada de forma direta tal situação não possui muita lógica, pois não é comum reduzir a disponibilidade e aumentar o OEE. Analisando mais detalhadamente os apontamentos de dados e também observando o processo em questão, é possível observar que as paradas operacionais estão mais voltadas a eventos de ajuste e limpeza de máquina, ou atividade de manutenção operacional do equipamento. Quando tais atividades são bem realizadas obtém-se um melhor desempenho da máquina, e também previne outros tipos de paradas involuntárias do equipamento, como por exemplo quebras, o que acarretariam em maiores perdas de disponibilidade. Conclui-se dessa forma, para esse processo em questão, que paradas operacionais auxiliam na preservação do equipamento e na melhoria de eficiência do mesmo.

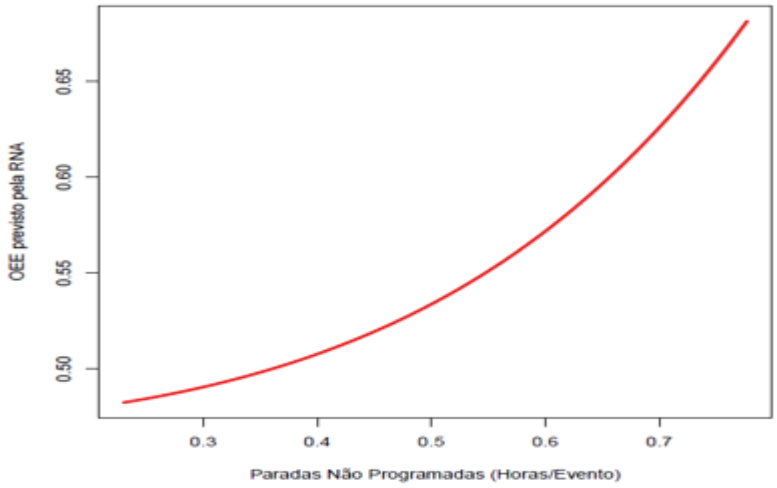
GRÁFICO 9 – OEE PREVISTO X PARADAS OPERACIONAIS



FONTE: O autor (2019).

Por outro lado, observando o GRÁFICO 10, as paradas não programadas também tenderam a um aumento do OEE. Tal fato necessita de uma melhor avaliação, sendo necessário realizar uma melhor análise, comparando a previsão do OEE, as paradas não programadas e outras variáveis. Esta análise foi realizada com auxílio de gráficos preditivos em 3D, e será apresentada no GRÁFICO 15.

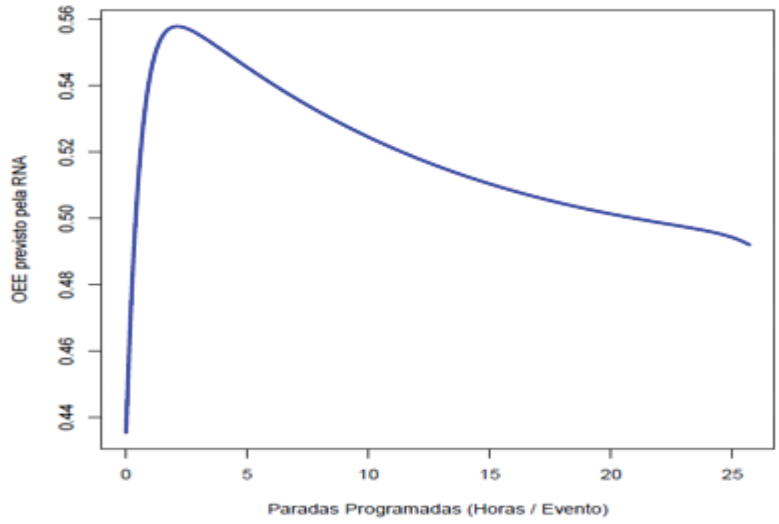
GRÁFICO 10 – OEE PREVISTO X PARADAS NÃO PROGRAMADAS



FONTE: O autor (2019).

No GRÁFICO 11 é demonstrada a influência das paradas programadas no OEE, sendo que tais paradas são aquelas que ocorrem em acordo com o PCP, sendo em muitos casos manutenções programadas, refeições, falta de programação de pedidos, entre outras. É possível observar que até um nível próximo de 5 horas existe um aumento da taxa de OEE, sendo reduzido significativamente com o aumento destas horas. Isso deve-se ao fato de que existem rotinas de paradas curtas de poucas horas para manutenção preventiva, as quais auxiliam na preservação do equipamento e consequente melhoria de eficiência. No caso de elevadas horas de máquina parada, no retorno da operação, o equipamento requer um maior tempo para inicialização e preparação de pré-produção, o qual acaba por influenciar na taxa de disponibilidade do OEE.

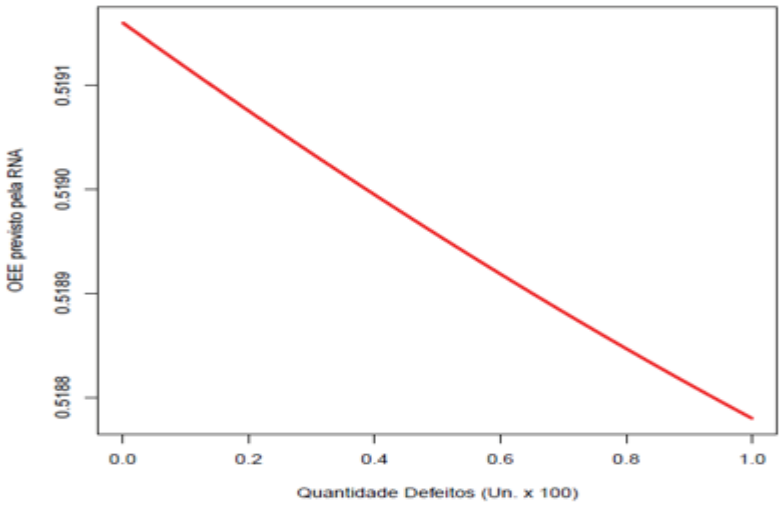
GRÁFICO 11 – OEE PREVISTO X PARADAS PROGRAMADAS



FONTE: O autor (2019).

O GRÁFICO 12 demonstra a relação entre a quantidade de defeitos e o OEE do equipamento. Nesse processo, como ocorreu com a taxa de produção e o tempo de produção, a quantidade de defeitos também não possui uma influência significativa no OEE, pois como as taxas de velocidade de máquina são elevadas e as produções ocorrem na grandeza de milhares de unidades por hora, os defeitos iriam impactar no OEE somente se existisse uma quantidade significativa de produtos não conforme.

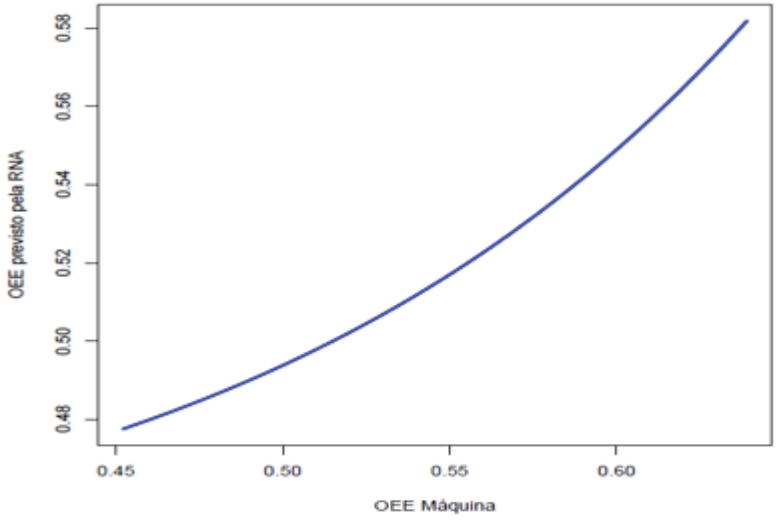
GRÁFICO 12 – OEE PREVISTO X QUANTIDADE DE DEFEITOS



FONTE: O autor (2019).

O GRÁFICO 13 demonstra que existe uma tendência de quanto melhor a eficiência de um produto em uma máquina, melhor a eficiência do equipamento. E também, quando um equipamento possui boa eficiência, também existe uma tendência de uma boa performance da fabricação do produto.

GRÁFICO 13 – OEE PREVISTO X OEE REALIZADO PELA MÁQUINA

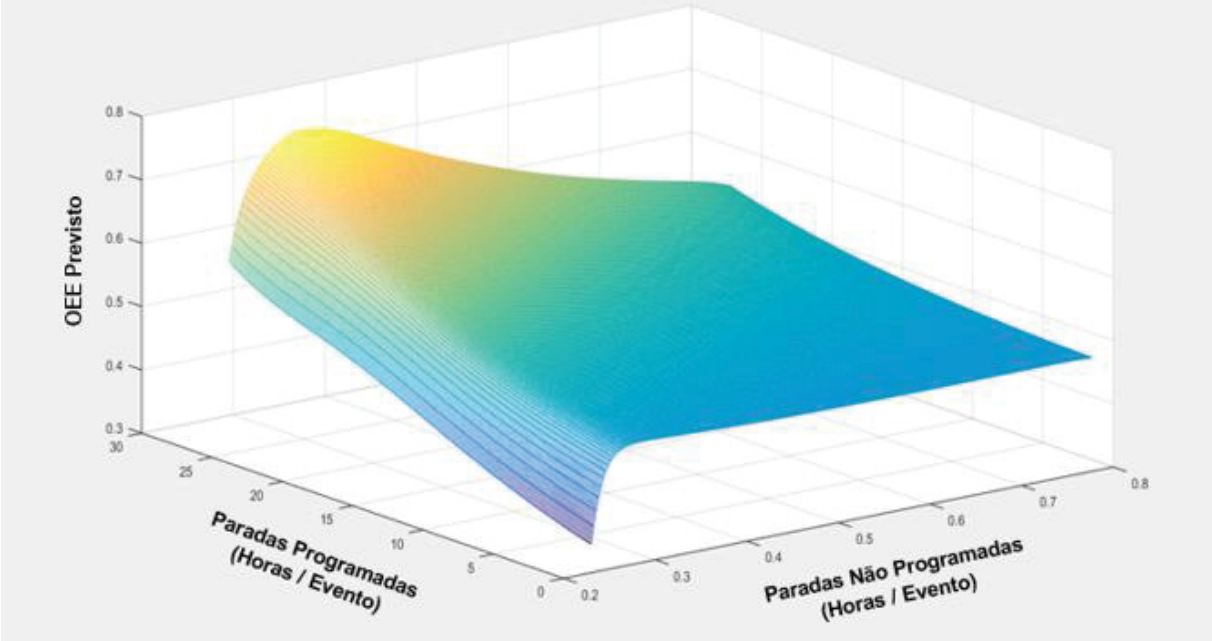


FONTE: O autor (2019).

Visando entender melhor alguns fatores de predição do OEE, foram realizadas comparações entre duas variáveis e o OEE previsto. Para ser possível tal análise foram criados os gráficos em 3D, os quais são apresentados nos gráficos de 14 a 20.

Analisando primeiramente a variável de “paradas não programadas”, pode-se observar pelo GRÁFICO 14 que na existência de paradas programadas, como manutenções preventivas, o índice de OEE aumenta com poucas horas de “paradas não programadas”, como panes involuntárias do equipamento.

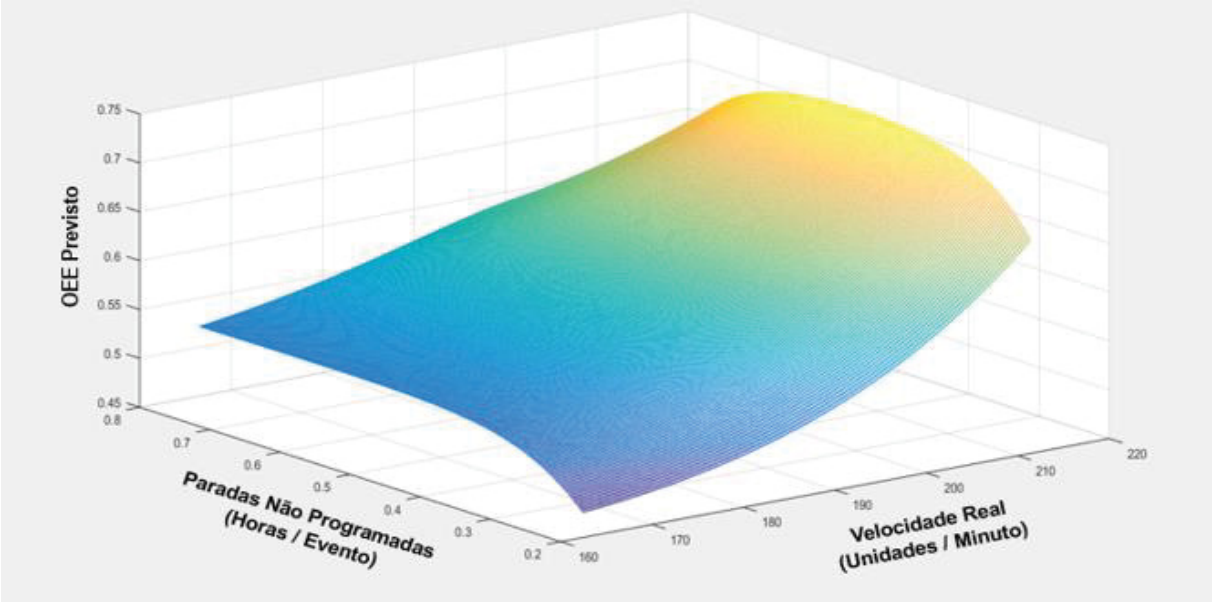
GRÁFICO 14 – OEE PREVISTO X PARADAS NÃO PROGRAMADAS X PARADAS PROGRAMADAS



FONTE: O autor (2019).

Reforçando-se mais o entendimento do comportamento das “paradas não programadas”, no GRÁFICO 15 é possível verificar que um aumento na taxa de velocidade tende a aumentar os índices de OEE, porém este mesmo aumento de velocidade gera uma quantidade maior de “paradas não programadas” com maiores valores de OEE. Ou seja, quando foi observado no GRÁFICO 10 que existia aumento de OEE pela elevação de “paradas não programadas”, na verdade existia um aumento de velocidade que propiciou o aumento de OEE por meio do aumento do índice de performance. O aumento de velocidade pode ter mascarado a perda de disponibilidade por “paradas não programadas”.

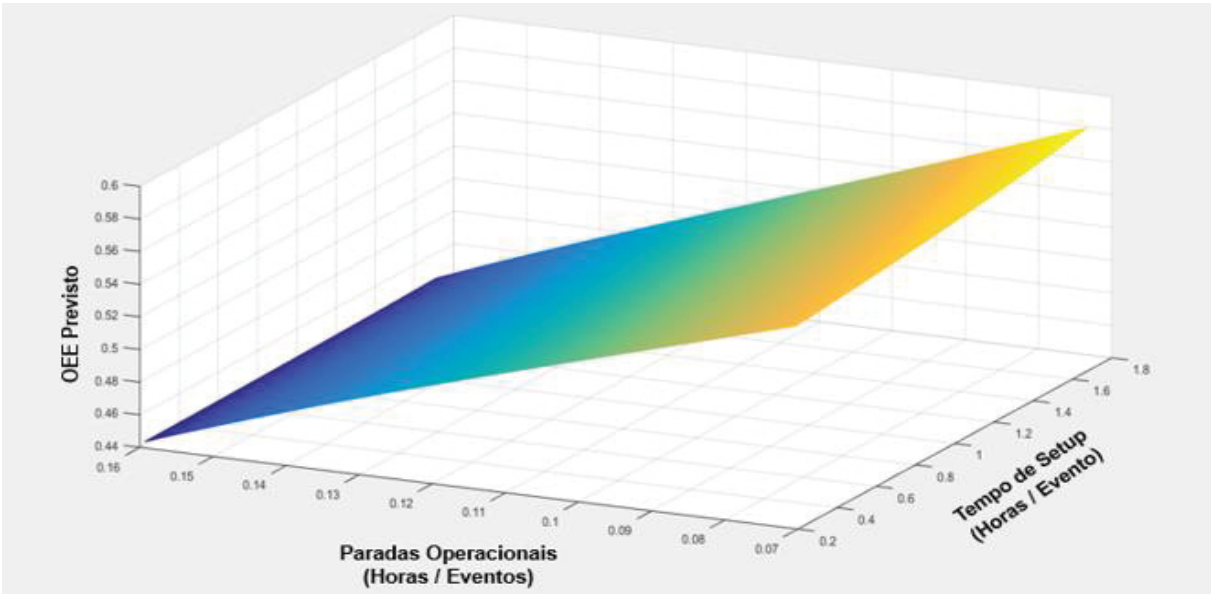
GRÁFICO 15 – OEE PREVISTO X VELOCIDADE REAL X PARADAS NÃO PROGRAMADAS



FONTE: O autor (2019).

No GRÁFICO 16 é possível observar que para maiores tempos ou eventos de *setup* podem existir índices mais elevado de OEE. Como já mencionado, tal fato poderia ser considerado como incoerente, pois como uma perda de disponibilidade por *setup* pode levar a maiores índices de eficiência. Mas neste caso, um melhor *setup* gerou uma maior estabilidade no equipamento, o que levou a uma redução de paradas operacionais, gerando aumento nas taxas de disponibilidade e consequentemente nos índices de OEE.

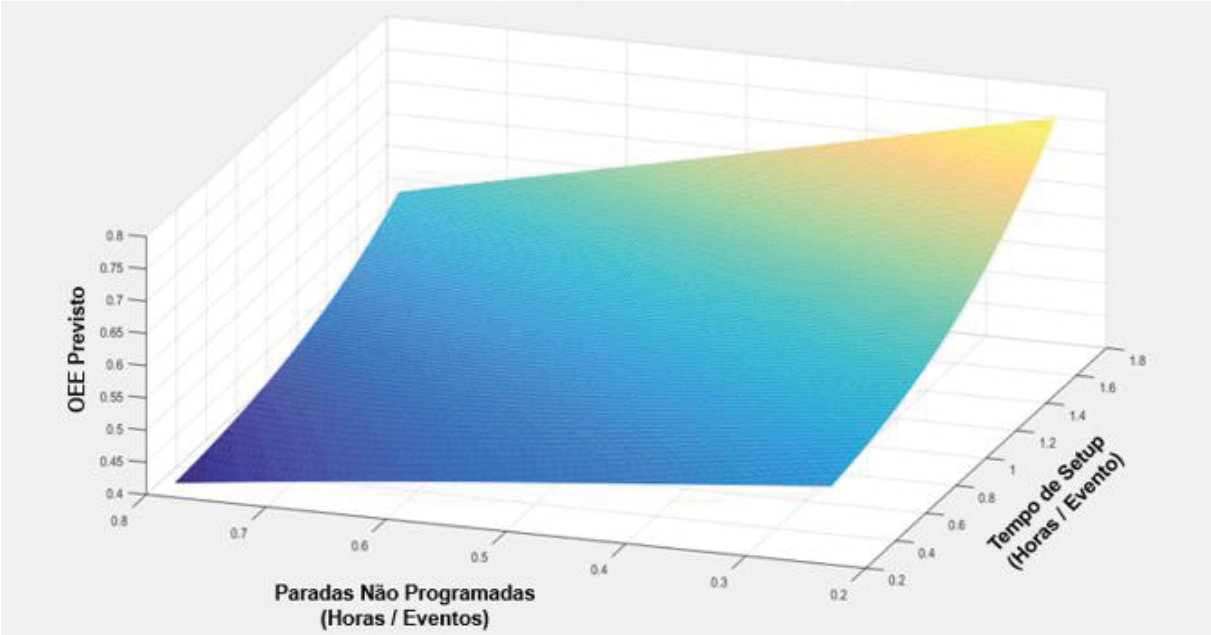
GRÁFICO 16 – OEE PREVISTO X TEMPO *SETUP* X PARADAS OPERACIONAIS



FONTE: O autor (2019).

No GRÁFICO 17 é possível avaliar que em casos de OEE com maiores índices, o tempo de *setup* pode ser maior como verificado no gráfico anterior. Porém quando existem baixos tempos de *setup*, além de contribuir negativamente devido ao aumento de paradas operacionais, também compromete ainda mais as taxas de disponibilidade pela tendência de contribuir com as paradas não programadas.

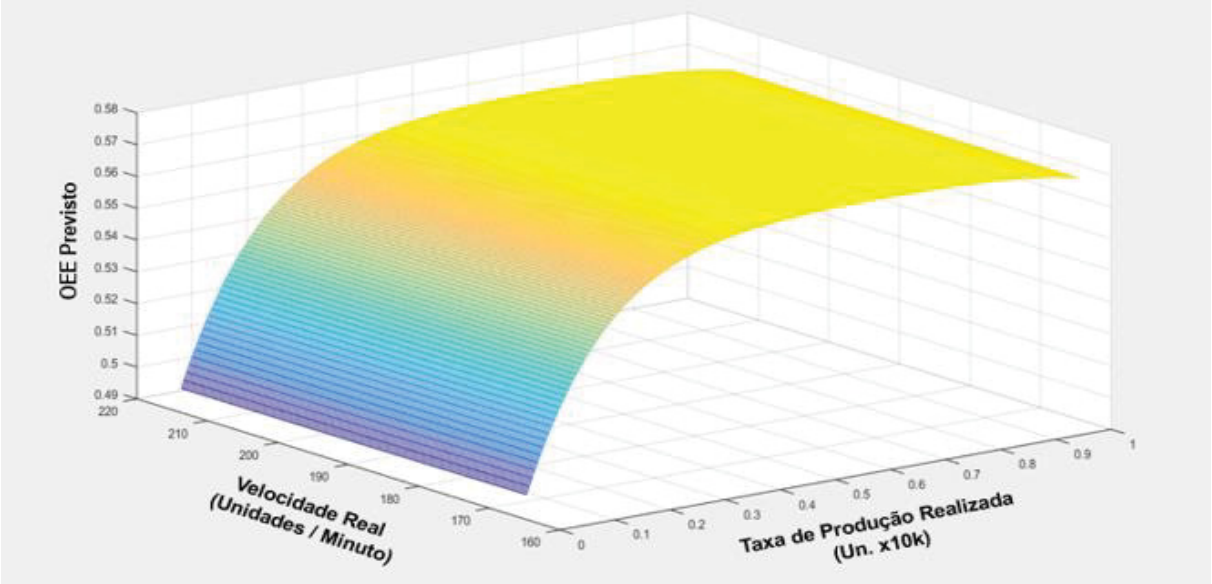
GRÁFICO 17 – OEE PREVISTO X TEMPO *SETUP* X PARADAS NÃO PROGRAMADAS



FONTE: O autor (2019).

O GRÁFICO 18 auxilia a evidenciar os índices de OEE pelas taxas de produção e de velocidade real do equipamento. Pode ser verificado que se as velocidades forem mais elevadas, porém sem uma maior taxa de produção o índice de OEE é menor. No caso de maiores velocidades, porém atrelado a maiores taxas de produção os índices de OEE se tornam mais elevados.

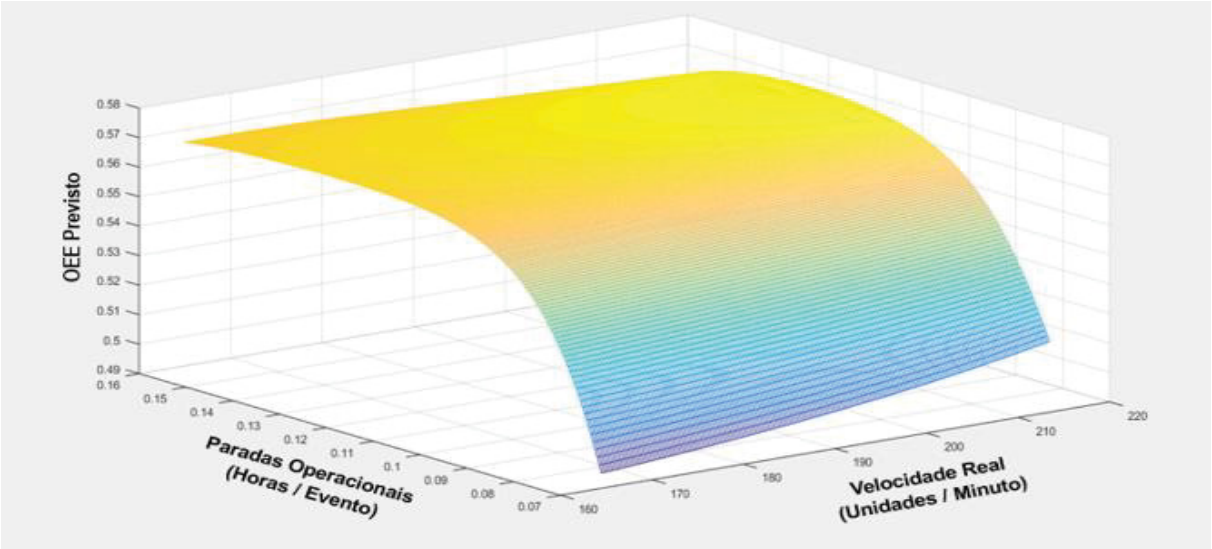
GRÁFICO 18 – OEE PREVISTO X TAXA DE PRODUÇÃO X VELOCIDADE REAL



FONTE: O autor (2019).

Como visto anteriormente, e observado no GRÁFICO 19, o aumento de velocidade leva a maiores índices de OEE, porém também pode levar a maiores tempos ou eventos de paradas operacionais, as quais podem representar maiores tempos de ajustes devido a perdas de posicionamento no equipamento pelo aumento de velocidade. Com o aumento de velocidade ocorre o aumento das taxas de produção, tal fato leva a um aumento de trocas frequênciadas no equipamento devido ao consumo de insumos ou materiais intercambiáveis. Tais paradas estão inclusas dentro das paradas operacionais.

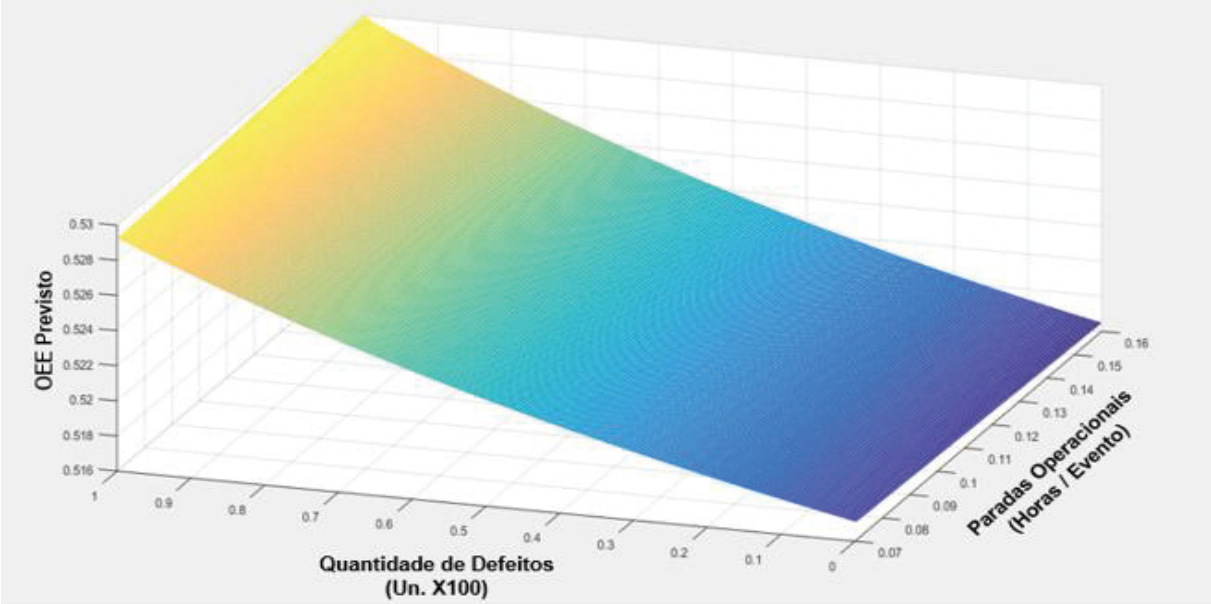
GRÁFICO 19 – OEE PREVISTO X VELOCIDADE REAL X PARADAS OPERACIONAIS



FONTE: O autor (2019).

Como visto anteriormente, para o tipo de processo estudado, a taxa de defeitos não possuem grande influência para o índice de eficiência do equipamento. No GRÁFICO 20, também é possível verificar que acontece o mesmo, em relação a taxa de defeitos e as paradas operacionais.

GRÁFICO 20 – OEE PREVISTO X PARADAS OPERACIONAIS X DEFEITOS



FONTE: O autor (2019).

5 CONCLUSÃO

O objetivo principal deste trabalho foi o de criar um modelo preditivo para análise de OEE baseado em um banco de dados gerado de forma automática pelo conceito da Internet das Coisas, tendo o intuito de aplicar métodos estatísticos e multivariados para a análise dos dados e criação de tal modelo.

Para obter conhecimentos específicos e mais aprofundados sobre esses processos foi realizada uma pesquisa bibliográfica sobre os temas da indústria 4.0, como *Big Data* e Internet das Coisas. Também foram pesquisados trabalhos relacionados a análises de predição utilizando técnicas estatísticas e multivariadas. Em relação ao OEE, sendo uma técnica utilizada em âmbito industrial, foram pesquisadas as formas de aplicação e utilização deste indicador.

Para o estudo foram consolidados um conjunto com milhares de dados de produção oriundos do processo de fabricação de embalagens de papel, representando três anos de fabricação, os quais continham referências sobre interferências de sazonalidade, mix de produção, influência dos equipamentos, estrutura de produtos, paradas de máquina, índices de defeitos, entre outros, os quais serviriam como base amostral para as análises a serem realizadas.

Na primeira fase de análise, relacionada ao tratamento e "limpeza" dos dados, foi necessário realizar a unificação de relatórios, eliminar informações e dados irrelevantes para a predição do OEE. Nestas atividades foi percebida grande dificuldade de desempenho computacional, pois na realização desta rotina estavam sendo manipuladas quantidades expressivas de dados. Como solução, iniciou-se a utilização do *software* R, o qual possui uma maior propensão em operar em ambientes como esse. Também foi utilizado um computador com maior desempenho (*workstation*). Entende-se que para próximos trabalho é necessário realizar uma avalição prévia de requisitos computacionais, relacionados a *hardwares* e *softwares* a serem aplicados.

Na primeira etapa do trabalho, aplicando a análise da matriz de correlação, demonstrou-se que é possível realizar uma análise de dados com quantidades parciais de uma amostra. Tal conclusão demonstra que não necessariamente é preciso utilizar todos os dados de uma população para retornar resultados significativos e satisfatórios. Além disso, quando é necessário trabalhar com grande

quantidade de dados, o requerimento computacional é mais exigido, tornando as análises mais morosas.

Para as etapas iniciais de construção do modelo, foi prevista a utilização de abordagens estatísticas multivariadas, por meio da aplicação de Análise de Componentes Principais e Regressão Linear Múltipla. Inicialmente tais técnicas apresentavam-se como as ideias para a criação do modelo, pois por meio da Análise de Componentes Principais poderiam ser reduzidas variáveis sem comprometer a análise, e utilizando a regressão linear objetivava-se criar uma equação que pudesse traduzir a predição do OEE por meio de inferências. Porém, durante a aplicação de tais técnicas percebeu-se que os resíduos referentes às equações de regressão estimadas não se ajustaram a uma distribuição normal, o qual é pré-requisito para a realização de inferências estatísticas sobre a mesma.

Entendeu-se que a utilização da técnica de regressão linear múltipla não funciona muito bem em ambientes com variáveis com comportamentos não-lineares. Em uma aplicação prática com grandes quantidades de dados apresentando comportamentos instáveis e muito variáveis, tal método não se apresentou como o mais adequado. Em virtude de tal conclusão optou-se por seguir com estudo, porém utilizando uma técnica multivariada que não precisasse do atendimento de pressupostos. Desta forma definiu-se utilizar a técnica de Redes Neurais Artificiais.

A utilização das RNAs trouxe uma nova forma de visualizar e analisar os dados. Sendo ela capaz de operar com dados multivariados, porém sem a necessidade dos mesmos possuírem um comportamento pré-estabelecido, ou seja, a relação entre as variáveis não requer comportamento linear. Este requisito foi identificado como um dos diferenciais na aplicação da Rede Neural Artificial.

Utilizando o algoritmo MLP, com a função de ativação sigmoide logística, foi possível utilizar dados de entrada que não dependiam de uma regra ou pressuposto, além de que, possui como característica a não dependência de relações pré-existentes dos dados de entrada da rede. Em algumas situações é exigida uma quantidade amostral significativa para a realização do treinamento e validação da rede, que neste caso foi atendido plenamente.

A aplicação da rede neural ocorreu em um conjunto de 10 variáveis pré-estabelecidas, sendo estas as possíveis variáveis preditoras para o OEE.

Como resultado da aplicação da rede neural nas variáveis, foram obtidos erros médios na ordem de 0,0159, os quais não ultrapassaram o valor de 3% pré-

estabelecido como erro máximo possível para a rede neural, validando assim a metodologia proposta. Além disso, ainda para análise dos resultados do modelo e observar o comportamento do OEE, foram fixadas algumas variáveis da rede, sendo variadas de uma em uma para geração de gráficos em duas dimensões (2D) e de duas em duas para geração de gráficos em três dimensões (3D). Pela fixação destas variáveis foi possível gerar cenários gráficos, os quais demonstravam o comportamento do OEE, pela magnitude da influência de cada uma das variáveis preditoras no OEE.

As análises gráficas demonstram o comportamento do OEE para cada tipo de variável. Em alguns casos trouxeram a confirmação a respeito de convicções empíricas do processo, como no caso do aumento do tempo de *setup*, o qual reduz a taxa de disponibilidade de máquina, a qual é diretamente proporcional a variação da taxa de OEE do equipamento.

Outros casos trouxeram a visualizações de situações desconhecidas, como a velocidade de máquina, a qual se comporta de forma ambígua em diferentes intervalos. Em um determinado intervalo, gera um OEE crescente, porém quando se aumenta mais a velocidade após certa faixa, o OEE tende a estabilizar e depois cair. Isso demonstra que o aumento de velocidade do equipamento por si só não garante uma eficiência do equipamento.

Outras situações demonstraram possibilidades de análises mais aprofundadas, com o auxílio de estratificação e análise pontual dos dados para um melhor entendimento. Porém de forma geral, a análise gráfica refletiu situações que estão presentes no processo, muitas vezes sendo conhecidas de forma empírica, porém não estando demonstradas de forma numérica ou explicita. Tal análise torna o entendimento acerca do desempenho, operação e eficiência dos equipamentos mais claros, precisos e de melhor entendimento para os participantes do processo.

Após as análises gráficas e considerando os erros médios gerados pela rede neural, foi possível afirmar que o modelo proposto atende os requisitos de predição do OEE, sendo tal modelo passível de replicação para demais cenários e processos. A aplicação das Redes Neurais Artificiais demonstrou-se como uma técnica adequada para ser utilizada em ambientes com grande quantidade e variabilidade de dados, além de ser eficaz para ambientes ondes tais dados não apresentam um comportamento normal e linear.

Assim conclui-se que o processo para criação de um modelo preditivo aplicado no indicador de OEE trouxe diversos aprendizados que agregaram conhecimentos teóricos, práticos e acadêmicos. A aplicação de vários métodos estatísticos e multivariados conduziu a um considerável aprofundamento sobre os temas, desde o processo de análises de dados, passando pela a interpretação dos resultados e testes, até chegar em conclusões acerca do modelo gerado. Em relação a aplicação prática, entende-se que tal modelo apresenta embasamento para aplicação em projetos futuros, além de possuir potencial para ser utilizado em ambientes industriais, auxiliando no planejamento e operação de uma fábrica por meio da predição do OEE de seus equipamentos e processos.

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Com a conclusão do trabalho, existem sugestões de aprofundamentos futuros, os quais podem ajudar na melhor compreensão do modelo desenvolvido e aperfeiçoar os métodos aplicados:

- Após o desenvolvimento da análise preditiva, é possível promover um estudo mais aplicado e prático, por meio de análises prescritivas.
- Para um melhor entendimento das variáveis independentes que compõem o OEE, é sugerido um desdobramento destas variáveis.
- O processo produtivo em questão, pode sofrer impactos em relação a sazonalidade, em virtude disso, estudo mais aprofundados sobre períodos específicos de produção podem ser realizados.
- Variações de mix de produtos podem ser uma característica a ser analisada mais detalhadamente, pois famílias de produto podem ter características diferenciadas e que impactam nos índices de OEE.
- Estudo mais aprofundados em relação aos equipamentos, analisando e verificando comportamentos de forma individual das máquinas.

REFERÊNCIAS

ASSEF, F. M. Algoritmos de classificação em aplicação financeira: avaliação de risco de crédito para pessoa jurídica. **Dissertação** - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação Engenharia de Produção. Curitiba. 2018.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2ª ed., Rio de Janeiro, LTC, 2011.

BOUROUCHE, J. M.; SAPORTA, G. **Análise de dados**. Tradução de Marcus Penchel. Rio de Janeiro: Zahar Editores, 1982. p.115. Tradução de: L´analyse des données.

DEZYRE. **Types of analytics: descriptive, predictive, prescriptive analytics**. Disponível em: < https://www.dezyre.com/article/types-of-analytics-descriptive-predictive-prescriptive-analytics/209>. Acesso em: 24 fev. 2018.

FLEURY, **A. Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. Rio de Janeiro: Elsevier: ABEPRO, 2012. p.32-46. ISBN 978-85-352-4850-0.

GAMESAUCE. **Predictive analytics in games**. Disponível em: <http://www.gamesauce.biz/2017/05/12/predictive-analytics-games/ >. Acesso em 24 fev. 2018.

GANDOMI, A.; HAIDER, M. **Beyond the hype: Big data concepts, methods, and analytics.** International Journal of Information Management, v. 35, n. 2, p. 137–144, 2014.

GARTNER. 2014. **BI: Analytics Moves To The Core.** Digital Business and Business Analytics – Timo Elliott's Blog. < https://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html>. Acesso em 10 mar 2018.

GARTNER. **IT Glossary**. Disponível em:<https://www.gartner.com/it-glossary/big-data>. Acesso em 17 fev 2018.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. ISBN 85-224-3169-8.

GOTTWALD, M. **Industria 4.0 e Internet Industrial:** Condições gerais, políticas e culturais. Ferdinand-Steinbeis-Institut. Steinbeis, 2016.

HAIR, J. F.; ANDERSON, R.E.; TATHAM R. L.; BLACK W. C. **Análise Multivariada de Dados**. Tradução de Adonai Schlup Sant'Anna e Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005. Tradução de: Multivariate Data Analysis. 593p.

HASHEM, I. A. T.; YAQOOB, I; ANUAR, N. B.; MOKHTAR, S.; GANI A.; KHAN S.U. **The rise of "Big Data" on cloud computing:** Review and open research issues. Information Systems, v. 47, p.98-115, 2015.

HASHIZUME, K.; ROSADO D. G.; FERNÁNDEZ-MEDINA E.; FERNANDEZ E. B. **Analysis of security issues for cloud computing**. Journal of Internet Services and Applications, 2013.

HAYKIN, S. **Redes neurais: princípios e práticas**. Tradução de Paulo Martins Engel, 2ª ed., Bookman, Porto Alegre, 2001.

HE, Q. P.; WANG, J. **Statistical process monitoring as a Big Data analytics tool for smart manufacturing**. Journal of Process Control, 2017.

IDC Analyze the Future. **THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East**. IDC Iview - EMC Corporation. 2012.

IDC Analyze the Future. **The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things**. EMC Digital Universe with Research & Analysis. 2014.

IBM. **Unidades de Medida para Dados de Armazenamento**. Disponível em: <https://www.ibm.com/support/knowledgecenter/pt-br/SSNE44_5.2.3/com.ibm.tpc_V523.doc/fqz0_r_units_measurement_data.html>. Acesso em: 17 mar. 2018.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6. ed. New Jersey: Prentice Hall, 1998. 773p.

JOLLIFFE, I. T. **Discarding Variables in a Principal Component Analysis**. I: Artificial Data. Journal of the Statistical Society. Series C (Applied Statistics). Vol. 21, No. 2, pp. 160-173, 1972.

JUNQUÉ DE FORTUNY, E.; MARTENS, D.; PROVOST, F. **Predictive Modeling with Big Data: *Is Bigger Really Better*?** Big Data, v. 1, n. 4, p. 215–226, 2013.

KENNEDY, R. K., **Understanding, Measuring and Improving Overall Equipment Effectiveness**: How to Use OEE to Drive Significant Process Improvement. CRC Press. Boca Raton, Florida, 2018.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: **International joint Conference on artificial intelligence**. [S.l.: s.n.], 1995. v. 14, p. 1137–1145.

LATTIN, J.; CARROL J. D.; GREEN, P. E.; **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011. 455p.

LEVINE, M. D.; BERENSON, L. M.; STEPHAN D., **Estatística: Teoria e Aplicações**. Rio de Janeiro: LTC – Livros Técnicos e Científicos, 1998.

LEE J.; KAO H.; YANG S. **Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment**. Procedia CIRP, v. 16, p. 3-8, 2014. ISSN 2212-8271.

LIBES, D.; SHIN, S.; WOO, J. **Considerations and recommendations for data availability for data analytics for manufacturing**. BT - 3rd IEEE International Conference on Big Data, IEEE Big Data 2015, October 29, 2015 - November 1, 2015. p. 68–75, 2015.

MANLY, B. J. F., **Métodos estatísticos multivariados: uma introdução.** Porto Alegre: Bookman, 2008. 229p.

MARKUS, F.; STEINBECK J., **Steigerung der Anlagenproduktivitat durch OEE – Management:** Definitionen, Vorgehen und Methoden – von manuell bis Industrie 4.0. Springer Gabler, doi: 10.1007/978-3-658-21456-2, ISSN 2197-6716. Wiesbaden, Germany, 2018.

MARQUESONE, R. **Big Data:** Técnicas e tecnologias para extração de valor dos dados. São Paulo: Editora Casa do Código, 2016.

MARTINS, R. A. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. Rio de Janeiro: Elsevier: ABEPRO, p. 47-63, 2012. ISBN 978-85-352-4850-0.

MARQUES, M. A. M. Aplicação da análise multivariada no estudo da infra-estrutura dos serviços de saúde dos municípios parananenses. **Dissertação** - Setor de Tecnologia e Setor de Ciencias Exatas, Programa de Pós-Graduação em Métodos Numéricos em Engenharia. Curitiba. 2005.

MARQUES, J. M. **Apostila de análise multivariada aplicada à pesquisa.** Universidade Federal do Paraná, Curitiba, PR, 2017.

MARQUES, J. M.; MARQUES M. A. M. **Estatística Básica para os Cursos de Engenharia**. Curitiba: Domínio do Saber, 2009.

MIGUEL, P. A. C.; SOUSA, R. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. ed. Rio de Janeiro: Elsevier: ABEPRO, p.131-147, 2012. ISBN 978-85-352-4850-0.

MIT Technology Review. **The Big Data Conundrum: How to Define It?** Disponível em: <https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/>. 2013. Acesso em 10 mar 2018.

MONTGOMERY D. C.; RUNGER G. C.; **Estatística Aplicada e Probabilidade para Engenheiros**. 4 ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A., 2009.

MONTGOMERY D. C.; PECK E. A.; VINING G. G.; **Introcuction to Linear Regression Analysis**. 5th ed. John Wiley & Sons, Inc. New Jersey, 2012.

MORIGGI, T. Predição de performance de cimentos compostos por meio da aplicação de redes neurais artificiais visando a garantia da qualidade. **Dissertação** - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação Engenharia de Produção. Curitiba. 2018.

NAKAJIMA, S. **Introduction to TPM: Total Productive Maintenance**. Originally published by the Japan Institute for Plant Maintenance. Tokyo, 1984. English translation copyright by Productivity Press. Portland, Oregon, 1988.

NAKANO, D. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. Rio de Janeiro: Elsevier: ABEPRO, p. 64-73, 2012. ISBN 978-85-352-4850-0.

OLSHANNIKOVA, E.; OMETOV A.; KOUCHERYAVY, Y; OLSSON, T. **Visualizing Big Data with augmented and virtual reality: challenges and research agenda**. Journal of Big Data, v. 2, n. 1, p. 1–27, 2015.

PHILIP CHEN, C. L.; ZHANG, C. Y. **Data-intensive applications, challenges, techniques and technologies: A survey on Big Data**. Information Sciences, v. 275, p. 314–347, 2014.

RASCHKA. **Esquema de validação *k-fold***. Disponível em: <https://sebastianraschka.com/blog/2016/model-evaluation-selection-part3.html> . Acesso em 11 mai. 2019.

SHAO, G.; SHIN, S. J.; JAIN, S. **Data analytics using simulation for smart manufacturing.** Proceedings - Winter Simulation Conference, v. 2015, n. December, p. 2192–2203, 2015.

SHIN, S. J.; WOO, J.; RACHURI, S. **Predictive analytics model for power consumption in manufacturing**. Procedia CIRP, v. 15, p. 153–158, 2014.

SIVARAJAH, U.; KAMAL M., IRANI Z., WEERAKKODY V. **Critical analysis of Big Data challenges and analytical methods.** Journal of Business Research, v. 70, p. 263–286, 2016.

SNIDERMAN, B.; MAHTO, M.; COTTELEER, M. **Industry 4.0 and manufacturing ecosystems:** Exploring the world of connected enterprises, 2016.

STAMATIS, D. H., **The OEE Primer: Understanding Overall Equipment Effectiveness, Reliability and Maintainability**. CRC Press. New York, 2010.

TALIA, D. **Clouds for scalable Big Data analytics**. Computer: IEEE Computer Society. v. 46, n. 5, p. 98–101, 2013.

WIKIPEDIA. **Gráfico de dispersão**. Disponível em: <https://pt.wikipedia.org/wiki/Gr%C3%A1fico_de_dispers%C3%A3o> . Acesso em 08 mar. 2018.

YIN, S.; KAYNAK, O. **Big data for modern industry:** Challenges and Trends. Preceedings of the IEEE, v.103, n.2, 2015.

ZICARI, R. V. **Big Data: Challenges and Opportunities**. Big Data computing, p. 103–128, 2014.

APÊNDICE A – ALGORITMOS UTILIZADOS NO *SOFTWARE* R

```
##### CONTAGEM DOS P-VALORES DE 712.000 OBSERVACOES (PASSO = 1.000)
EVENTOS_PVAL<-1
obs_AL<-1000
t0<-Sys.time()
for(i in 1:712){
  PV_obs_AL<-BD[sample(1:nrow(BD),obs_AL,replace = F),]
  cors_pv_AL<-rcorr(as.matrix(PV_obs_AL),type = "spearman")
  PV_AL<-cors_pv_AL$P
  EVENTOS_AL<-length(which(PV_AL<0.05))
  EVENTOS_PVAL<-c(EVENTOS_PVAL,EVENTOS_AL) #cbind
  obs_AL<-obs_AL+1000
  cat(i,'\n')}
t1<-Sys.time()
cat('Tempo total aleatório=',t1-t0,'\n')

##### CORRELACAO #####
# install.packages("corrplot", repos="http://cran.rstudio.com/", dependencies=TRUE)
# install.packages("Hmisc", repos="http://cran.rstudio.com/", dependencies=TRUE)
# install.packages("PerformanceAnalytics", repos="http://cran.rstudio.com/", dependencies=TRUE)
M14_CORR<-cor(M14_N, method="spearman") # Correlacao de Spearman
# cor.test(M14_N, method="spearman") # Teste estatistico correlacao Spearman
# M14_CORR_HIST<-M14_CORR[,1:21] # Escolher a quantidade de colunas
M14_CORR_NP<-rcorr(as.matrix(M14_N),type = "spearman")
corrplot(M14_CORR_NP$r,p.mat = M14_CORR_NP$P,sig.level = 0.05,method = "number",type =
"lower")

##### ANALISE COMPONENTES PRINCIPAIS #####
M14_SY<-M14_N[,-1] # ELIMINANDO A VARIAVEL RESPOSTA OEE (Y)
M14_SV<-M14_SY[,c(-1,-7)] # ELIMINANDO AS VARIAVEIS SEM CORRELACAO
M14_CORR_CP<-cor(M14_SV, method="spearman") # Correlacao de Spearman
M14_CORR_CP_NP<-rcorr(as.matrix(M14_SV),type = "spearman")
corrplot(M14_CORR_CP_NP$r,p.mat = M14_CORR_CP_NP$P,sig.level = 0.05,method =
"number",type = "lower")
M14_CORR_AV<-eigen(M14_CORR_CP_NP$r)
M14_CORR_AVal<-M14_CORR_AV$values
M14_CORR_AVet<-M14_CORR_AV$vectors
barplot(M14_CORR_AVal)
```

```
##### REGRESSAO LINEAR #####
# COM INTERCEPT
M14_RL_B2 <-
lm(OEE~Alt_F+Esc+H_Prd+H_Stp+Larg_F+OEE_Maq+Par_Prog+T_Valv,data=M14_RL_B2)
summary(M14_RL_B2)
M14_RL_B4 <-
lm(OEE~Alt_A+Esc+H_Stp+OEE_Maq+Par_NProg+Peso+T_Corte+Vel_Teo,data=M14_RL_B4)
summary(M14_RL_B4)
# SEM INTERCEPT
BD_M14_RL_B2<-readXL("C:/Users/Dayub/Desktop/04-Relatórios Dados/Analise_19-03-
10_Nova/M14/RL/M14_20K_AL_RL.xlsx",rownames=FALSE, header=TRUE, na="", sheet="J_B2",
stringsAsFactors=TRUE)
BD_M14_RL_B4<-readXL("C:/Users/Dayub/Desktop/04-Relatórios Dados/Analise_19-03-
10_Nova/M14/RL/M14_20K_AL_RL.xlsx",rownames=FALSE, header=TRUE, na="", sheet="J_B4",
stringsAsFactors=TRUE)
M14_RL_B2_SI <-
lm(OEE~0+H_Stp+Par_Op+Par_Prog+OEE_Maq+Gram+T_Corte+T_Fundo+Esc,data=BD_M14_RL_
B2)
summary(M14_RL_B2_SI)
M14_RL_B4_SI <-
lm(OEE~0+H_Stp+Par_Op+Par_Prog+OEE_Maq+Alt_A+Gram+T_Fundo+Esc,data=BD_M14_RL_B4
summary(M14_RL_B4_SI)

##### REDES NEURAIS ARTIFICIAIS #####
# --- PRODUTO A ---- #
n<-nrow(dadosA)
resultadoA<-data.frame(semente=NULL,n.neuronios=NULL,erro=NULL,tempo=NULL)
for (k in 1:15){
  set.seed(k)
  t<-sample((1:n),ceiling(n*0.8))
  set.seed(k)
  v<-sample((1:n)[-t],floor(n*0.2))
  treina<-dadosA[t,]
  valida<-dadosA[v,]
  for(i in 1:20){
  ti<-Sys.time()
  set.seed(k)
  rede<-
neuralnet(OEE~Prod+V_Real+H_Prod+H_Stp+Par_Op.+Par_N.Prog.+Par_Prog.+Def+Vel_Teo+OEE
_Maq,data=treina,hidden=i)
```

```
    tf<-Sys.time()
    if(length(rede$net.result)>0){
    previsao<-as.vector(compute(rede,valida[,2:11])$net.result)
    erro<-sqrt((sum(valida[,1]-previsao)^2)/length(previsao))
    } else {
      erro<-NA    }
    resultadoA<-rbind(resultadoA,data.frame(semente=k,n.neuronios=i,erro=erro,tempo=tf-ti))
    cat('semente = ', k, '    n.neuronios = ',i,'    erro = ',erro, '   tempo = ',tf-ti,'\n')  } }
# ---- Escolha do número de neurônios na camada oculta pelo menor erro médio  -------- #
media_erros_prodA<-NULL
media_erros_prodB<-NULL
for (i in 1:20){
  media_erros_prodA<-c(media_erros_prodA,mean(resultadoA[!is.na(resultadoA$erro) &
resultadoA$n.neuronios==i,'erro']))
  media_erros_prodB<-c(media_erros_prodB,mean(resultadoB[!is.na(resultadoB$erro) &
resultadoB$n.neuronios==i,'erro'])) }
n.neuronios_prodA<-which.min(media_erros_prodA)
n.neuronios_prodB<-which.min(media_erros_prodB)
# ---- Teste do k-fold  ---- #
# produto A
n<-nrow(dadosA)
# embaralhando os dados
set.seed(1)
dadosA<-dadosA[sample((1:nrow(dadosA)),nrow(dadosA)),]
t1<-1:14120; v1<-14121:17650
t2<-3531:17650; v2<-1:3530
t3<-c(1:3530,7061:17650); v3<-3531:7060
t4<-c(1:7060,10591:17650); v4<-7061:10590
t5<-c(1:10590,14121:17650); v5<-10591:14120
listat<-list(t1,t2,t3,t4,t5)
listav<-list(v1,v2,v3,v4,v5)
erroA<-NULL;sementeA<-NULL
for (k in 1:5){
    treina<-dadosA[listat[[k]],]
  valida<-dadosA[listav[[k]],]
  i<-1; flag<-TRUE
  while(flag){
    set.seed(i)
```

```
   rede<-
neuralnet(OEE~Prod+V_Real+H_Prod+H_Stp+Par_Op.+Par_N.Prog.+Par_Prog.+Def+Vel_Teo+OEE
_Maq,data=treina,hidden=n.neuronios_prodA)
   if(length(rede$net.result)>0){
     flag<-FALSE
   } else i<-i+1      }
  previsao<-as.vector(compute(rede,valida[,2:11])$net.result)
  erroA<-c(erroA,sqrt((sum(valida[,1]-previsao)^2)/length(previsao)))
  sementeA<-c(sementeA,i)}
validacaoA<-data.frame(erro=erroA,semente=sementeA)
# ------- REDE FINAL  ------ #
#  produto A
set.seed(1)
rede1<-
neuralnet(OEE~Prod+V_Real+H_Prod+H_Stp+Par_Op.+Par_N.Prog.+Par_Prog.+Def+Vel_Teo+OEE
_Maq,data=dadosA,hidden=n.neuronios_prodA)

##### GRAFICOS RNA #####
#---- graficos 2D ----#
dados<-dados1
rede<-rede1
tam<-1000
pdf('variacoes_prod1.pdf') # aqui vc está gerando um pdf com todos os gráficos
#Prod
dados_simul<-
data.frame(Prod=seq(min(dados$Prod),max(dados$Prod),len=tam),V_Real=rep(median(dados$V_Re
al),times=tam),H_Prod=rep(median(dados$H_Prod),times=tam),H_Stp=rep(median(dados$H_Stp),tim
es=tam),Par_Op.=rep(median(dados$Par_Op.),times=tam),Par_N.Prog.=rep(median(dados$Par_N.Pr
og.),times=tam),Par_Prog.=rep(median(dados$Par_Prog.),times=tam),Def=rep(median(dados$Def),ti
mes=tam),Vel_Teo=rep(median(dados$Vel_Teo),times=tam),OEE_Maq=rep(median(dados$OEE_Ma
q),times=tam))
simul<-as.vector(compute(rede,dados_simul)$net.result)
plot(seq(min(dados$Prod),max(dados$Prod),len=tam),simul,type='l',main='OEE Previsto x Taxa
Produção Realizada',xlab="Produção Realizada",ylab="OEE previsto pela RNA")

#---- graficos 3D ----#
dados<-dados1
rede<-rede1
tam<-200 # não mudar esse valor, pois o gráfico em 3D fica muito pesado com uma sequencia maior
#Prod e V_Real
```

```
d<-
data.frame(Prod=seq(min(dados$Prod),max(dados$Prod),len=tam),V_Real=seq(min(dados$V_Real),
max(dados$V_Real),len=tam),H_Prod=rep(median(dados$H_Prod),times=tam),H_Stp=rep(median(d
ados$H_Stp),times=tam),Par_Op.=rep(median(dados$Par_Op.),times=tam),Par_N.Prog.=rep(median
(dados$Par_N.Prog.),times=tam),Par_Prog.=rep(median(dados$Par_Prog.),times=tam),Def=rep(medi
an(dados$Def),times=tam),Vel_Teo=rep(median(dados$Vel_Teo),times=tam),OEE_Maq=rep(median(
dados$OEE_Maq),times=tam))
Prod<-seq(min(dados$Prod),max(dados$Prod),len=tam)
V_Real<-seq(min(dados$V_Real),max(dados$V_Real),len=tam)
z<-outer(Prod,V_Real, function(a,b)
{return(as.vector(compute(rede,data.frame(Prod=a,V_Real=b,H_Prod=rep(median(dados$H_Prod),ti
mes=tam),H_Stp=rep(median(dados$H_Stp),times=tam),Par_Op.=rep(median(dados$Par_Op.),times
=tam),Par_N.Prog.=rep(median(dados$Par_N.Prog.),times=tam),Par_Prog.=rep(median(dados$Par_
Prog.),times=tam),Def=rep(median(dados$Def),times=tam),Vel_Teo=rep(median(dados$Vel_Teo),tim
es=tam),OEE_Maq=rep(median(dados$OEE_Maq),times=tam)))$net.result))} )
#persp(Prod,V_Real,z)
write.xlsx(Prod,'Prod&V_Real.xlsx','Plan1',row.names=F,col.names=F)
write.xlsx(V_Real,'Prod&V_Real.xlsx','Plan2',row.names=F,col.names=F,append=T)
write.xlsx(z,'Prod&V_Real.xlsx','Plan3',row.names=F,col.names=F,append=T)
# esse código está variando as variáveis Prod e V_Real, e gravando uma planilha excel para exportar
para o Matlab e fazer o gráfico em 3D.
```